3-30-2015

Multispectral Image Road Extraction Based Upon Automated Map Conflation

Bin Chen

Multispectral Image Road Extraction Based Upon Automated Map Conflation

by

Bin Chen

B.S. Nanjing University of Science and Technology, 2006
M.S. Nanjing University of Science and Technology, 2008

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Chester F. Carlson Center for Imaging Science
College of Science
Rochester Institute of Technology

March 30, 2015

Signature of the Author _____

Accepted by _____
Coordinator, Ph.D. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Bin Chen
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation required for the
Ph.D. degree in Imaging Science

_____
Dr. Anthony Vodacek, Dissertation Advisor

_____
Dr. Nathan D. Cahill

_____
Dr. John P. Kerekes

_____
Dr. Pengcheng Shi

_____
Date

Multispectral Image Road Extraction Based Upon Automated Map Conflation

by

Bin Chen

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

Road network extraction from remotely sensed imagery enables many important and diverse applications such as vehicle tracking, drone navigation, and intelligent transportation studies. There are, however, a number of challenges to road detection from an image. Road pavement material, width, direction, and topology vary across a scene. Complete or partial occlusions caused by nearby buildings, trees, and the shadows cast by them, make maintaining road connectivity difficult. The problems posed by occlusions are exacerbated with the increasing use of oblique imagery from aerial and satellite platforms. Further, common objects such as rooftops and parking lots are made of materials similar or identical to road pavements. This problem of common materials is a classic case of a single land cover material existing for different land use scenarios.

This work addresses these problems in road extraction from geo-referenced imagery by leveraging the OpenStreetMap digital road map to guide image-based road extraction. The crowd-sourced cartography has the advantages of worldwide coverage that is constantly updated. The derived road vectors follow only roads and so can serve to guide image-based road extraction with minimal confusion from occlusions and changes in road material. On the other hand,

the vector road map has no information on road widths and misalignments between the vector map and the geo-referenced image are small but nonsystematic. Properly correcting misalignment between two geospatial datasets, also known as map conflation, is an essential step.

A generic framework requiring minimal human intervention is described for multispectral image road extraction and automatic road map conflation. The approach relies on the road feature generation of a binary mask and a corresponding curvilinear image. A method for generating the binary road mask from the image by applying a spectral measure is presented. The spectral measure, called anisotropy-tunable distance (ATD), differs from conventional measures and is created to account for both changes of spectral direction and spectral magnitude in a unified fashion. The ATD measure is particularly suitable for differentiating urban targets such as roads and building rooftops. The curvilinear image provides estimates of the width and orientation of potential road segments. Road vectors derived from OpenStreetMap are then conflated to image road features by applying junction matching and intermediate point matching, followed by refinement with mean-shift clustering and morphological processing to produce a road mask with piecewise width estimates.

The proposed approach is tested on a set of challenging, large, and diverse image data sets and the performance accuracy is assessed. The method is effective for road detection and width estimation of roads, even in challenging scenarios when extensive occlusion occurs.

Acknowledgements

I am deeply grateful to my advisor Dr. Anthony Vodacek, for his constant guidance and encouragement. Over the years, Tony has always been supportive and open-minded. Under his direction, I have enjoyed the freedom and pleasure of pursing varied research interests.

I also want to thank my dissertation committee members. Dr. Nathan Cahill is always willing to share his knowledge and to provide insightful suggestions on my research. I am indebted to Dr. John Kerekes and Dr. Pengcheng Shi for taking time to serve on my committee and providing valuable feedback on my dissertation.

I would like to express special appreciation to Dr. Weihua Sun. Weihua is a wonderful friend and also an excellent teammate to work with. The collaboration with him is always pleasant and rewarding. My research would not have been smooth without his selfless help.

My thank also goes to Susan Chan and Cindy Schultz for helping me stay on the right track and keeping things well organized.

I would like to extend my thanks to all fellow students in the Chester F. Carlson Center for Imaging Science. I have really enjoyed the time with my officemates: Dr. Lingfei Meng, Dr. Jiashu Zhang, Dr. Kelly Canham, and Dr. Sanjit Maitra. Also special thank goes to Yang Hu, Yilong Liang, and Wei Yao for the help in various ways.

I wish to recognize the generous financial support from the Chester F. Carlson Center for Imaging Science, NSF (National Science Foundation), and AFOSR (Air Force Office of Scientific Research) DDDAS (Dynamic Data Driven Applications Systems) program.

I would also like to thank DigitalGlobe Inc. for generously providing WorldView-2 and GeoEye-1 satellite imagery for this research.

Most importantly, I devote my utmost gratitude to my parents and my fiancée Jiexia for their unwavering love and care. This dissertation would not have been possible without their unconditional support and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Road network extraction from remotely sensed imagery has been actively explored for a very long time. Such keen interest among researchers is primarily aroused by its potential important applications. Roads are a high-value object class in an image captured by a space-borne imaging sensor. Enormous efforts have been invested over decades on comprehending the geometric structures and spectral footprints of roads either in a supervised or non-supervised fashion. Unfortunately, the performance of existing techniques is quite limited due to a series of challenges, which will be elaborated in the following section. Meanwhile, there have been few reported automatic road extraction solutions that are universally accurate and robust on different kinds of image scenes. Manual labeling of roads could be a more accurate way but is also notoriously time-consuming and labeling consistency can sometimes be a problem.

A related topic called map-to-imagery conflation is a relatively new research area, but has only seen success on a limited number of images. The objective of conflation is to fuse digital road map and geo-referenced imagery, which are two distinctive types of geospatial data and are only linked by their geographic coordinates, together. This work is motivated by the demand of correcting the inevitable misalignment between heterogeneous data sources and then presenting richer geospatial information in a combined framework. Conflation appeals to

Figure 1.1: *Challenges in image-based road extraction.*

researchers when online digital map data and remotely sensed imagery become more easily accessible, e.g., Google maps, and location based service (LBS) gains in popularity with the rise of mobile devices. Map conflation, however, is often accomplished by human operators, which also requires a large amount of labor and they may struggle to keep up with the frequent data update.

It is the goal of gaining a better characterization of unique road features and also heading towards a more optimal design of an automated and robust system for map conflation and road extraction that motivates the research presented in this dissertation.

1.1 Background and Objective

Image-based road network extraction enables many important and diverse applications, such as geographic information system (GIS) database update [4], urban mapping and planning [5], drone navigation [6], and intelligent transportation [7]. There are, however, a number of challenges in road detection from an image. Referring to Fig. 1.1, road pavement material, width, direction and even topology vary across a single image. Complex junctions add to the difficulties of the extraction problem. Complete or partial occlusions caused by nearby buildings, trees and the shadows cast by them make maintaining road connectivity difficult. Further, in a high-resolution image vehicles or even lane marks, are visible

and create an unpredictable spatial and spectral texture. Another challenge is related to the fact that common objects such as rooftops and parking lots that are made of materials similar or identical to road pavement, e.g., asphalt or concrete. They are often spatially connected or adjacent to roads. This added complexity weakens the separability of road pavement from background objects. Previous work has applied advanced techniques to address the road extraction problem, but the extracted road network always has spurious or missing road branches and its quality depends on scene content. Consequently, these methods almost inevitably face the limitation that the connectivity of road network is not maintained due to fragmentation of extracted road pieces. In addition, normally only road centerlines are extracted from the image, and road width is not routinely recovered. To effectively address this problem, incorporation of additional prior road information becomes imperative.

Introduction of an existing digital road map is an important step towards robust road network extraction since it greatly enriches our knowledge about the roads in a geo-referenced image. The road map is an informative data source that can be used as the guide for road extraction because it provides examined information about road skeleton shape and topology. More importantly, the hint of a road presence in an image is already extremely useful when compared with the "blind search" of purely image-based approaches that do not have awareness of such information. However, misalignment between the two datasets - the vector map and the raster image - are persistent and correction is necessary. This process of matching two sources of geographic data is referred to as spatial data conflation. Used in the community of GIS, *conflation* is defined as a process to combine cartographic data together with another geospatial data source to yield a superior maps with better quality and enhanced interpretability. The other data source could be another map, an image, or other relevant data. Spatial inconsistency or misalignment always exists for multi-source data when they are matched and fused, and the purpose of conflation is to co-register these different data sources. Only one type of map conflation, named vector-to-imagery conflation, is discussed in this research. Note an image will be used as the reference, to

which vector road data will be conflated. Following conflation, the map features, are used to guide extraction of road pixels and to generate a binary mask labeled with road pixels.

In this research, we present a novel system that is capable of extracting road pixels in a multispectral image and simultaneously conflating a vector road map to the geo-referenced imagery. It uses a digital road map from OpenStreetMap to perform map conflation to a geo-referenced image and guide image-based road extraction based on unique road features.

The general objective of this dissertation is to establish a generic road extraction algorithm pipeline that fits in various types of image scenes and the itemized objectives are listed as follows:

- Exploit representative road features that can be applied to unambiguously determine the presence and characteristic of the road network.
- Develop a vector road map to raster image conflation algorithm based on the extracted road features.
- Implement an image-based road network extraction system following automated map conflation.

1.2 Related Work

Image-based road extraction has been extensively explored in remote sensing community for over four decades. [8] With the advent of high spatial resolution multispectral sensors and ubiquitous geospatial applications, road network detection continues to be an active research topic. A summary of previous work on image road extraction is organized according to the use of prior data. Since there are hundreds of research papers on image road extraction, only an overview of recent related work is presented.

Figure 1.2: *Issues with road extraction with no prior data. [1] Circles indicate miss or false detections.*

1.2.1 Road extraction with no prior data

Road extraction with no prior data has been explored ever since the emergence of this research topic. [9] gave an overview and categorization of the related work up until 2007. We will give a brief literature review of those works after that. In [10], Hu et al. used hierarchical grouping strategy to automatically extract main road centerlines. Hu et al. [9] employed spoke wheel operator and toe-finding algorithms to track road footprints, followed by refinement by road tree pruning. A multistage framework for automatically extracting roads from high-resolution images based on salient features was introduced in [11]. Lin et al. [12] presented a semi-automatic approach to extract roads by finding lane markings and using interlaced template matching. A novel system involving three modules - probabilistic road center detection, road shape extraction, and graph-theory-based road network formation - is proposed in [13]. In [14], a CRF model was developed for road labeling, in which the prior is represented by higher-order cliques. Based on the assumption of color uniformity of the roads, Chai et al. [1] proposed a stochastic model for extracting line-networks from images by applying junction-point processes. However, plenty of miss and false detections are visible in their results as shown in Fig. 1.2. Sun and Messinger [15] proposed a knowledge-based automated road network extraction system based on extracted curvilinear

structures. Shi et al. [16] presented an integrated framework for road centerline extraction, which includes methods like mathematical morphology, locally weighted regression, and tensor voting.

The common problem of all the approaches is that topographical completeness, i.e., road connectivity, cannot be guaranteed. Due to the inevitable occlusions of buildings, tree canopies, and vehicles on the roads, extracted but broken roads always exist and that hinders their practical applications. Moreover, these methods often fail to extend to large-scale and complex image scenes, where road features change, illustrating the challenge of image-based road extraction without any prior information about the roads. The difficulty of directly extracting a road network from an image prompts researchers to seek solutions beyond the image itself.

1.2.2 Road extraction using prior data

To address the aforementioned problems, additional data sources are fused into a hierarchical framework and serve as the guidance for road extraction. For example, LiDAR (Light Detection And Ranging) or DSM (digital surface model) data are often used as a reliable spatial prior to assist road extraction, which leads to a desired separation of ground and above-ground objects [17, 18, 19, 20]. However, such data are not readily available and often require registration and region extraction before they can be used to match with road features in the image.

Another auxiliary data set more pertinent to road extraction is the geo-referenced digital road map, which is typically in vector graph format with vertices (or node points) and corresponding edges linking them. There are occasionally attributes associated with each vertex or edge describing road type, name, width, speed limit, etc., though these metadata are not always available. Further, imperfections occur when the road map is geo-registered to the image: spatial misalignment is common due to a variety of errors. The misalignment cannot be easily rectified via linear transforms, e.g., translation and rotation, due to non-systematic displacement. Map conflation is needed to correct errors and improve data quality.

Therefore, vector-to-imagery conflation or vice versa is being actively explored and is one of the challenging problems in GIS. Although the purpose of road map conflation is not originally designed for road extraction, in fact conflated road vectors correspond to road centerlines in the image and can be useful to extract road pixels.

Only those works most relevant to ours will be presented as follows. Readers can refer to [21] for a thorough review of the historical research on map conflation over the past few decades. Zhang [17] proposed an automatic system for updating outdated roads by combining multiple features and cues derived from various data sources, including a road database. Baltsavias [22] gave an overview of image-based building and road extraction systems using existing geodata. In [23], Chen et al. described automatic multi-source vector to imagery conflation approach by performing localized image processing on the orthoimagery. [24] proposed an automatic large-scale approach to align a raster image to Google maps by computing local translation within each image tile, followed by global thin-plate-spline warping based on tile control point pairs and confidence values. But the inconsistent misalignment within each tile cannot be corrected. Peng et al. [25] incorporated a GIS map of the road network as prior energy into a phase field higher order active contour model. Song et al. [21] used a spatial length-width contextual measure to define a binary road mask, which was later used in intersection and termination extraction. Matched point pairs were used as control points to perform rubber-sheeting transform, followed by a modified snake algorithm, which, however, relies on the creation of a skeleton-style binary road mask [26], to move intermediate vector road points towards the binary image road mask. In [27], Mnih and Hinton presented an approach for automatically detecting roads in aerial imagery using neural networks, whose training data is derived from a rasterized vector road map already aligned with that image. Zhang et al. [28] proposed a method for conflation of road network data and a geo-referenced image by using a sparse matching algorithm to find the correspondence of extracted road features. Lu et al. [29] introduced an automatic method for the estimation of presumed affine transformation parameters between vector

(a) *Chen et al. 2006 [23]* **(b)** *Wu et al. 2007 [24]* **(c)** *Song et al. 2009 [21]*

Figure 1.3: *Related work in map conflation.*

road maps and images. Vector maps are used as guidance for feature extraction. Yuan and Cheriyadat [30] presented a method that accurately segments road regions with a weak supervision provided by OpenStreetMap data and using factorization-based segmentation algorithm.

Three representative conflation methods [21, 23, 24] are selected to provide an intuitive perception of their performances. It can be easily seen in Fig. 1.3 that various errors still exist in their final results. One of the limitations associated with the above methods is that they usually extract road centerlines without any width estimate that should be an integral part of the extracted road information. Though road widths can be obtained directly from the map data [23, 24], its availability cannot be ensured. In addition, a rubber-sheeting transform or thin-plate-spline transform [21, 23, 24, 28] is normally used to conflate road segments based on paired control point. However, if no further processing step is involved, the residual misalignment error will not be eliminated. Furthermore, most of the existing approaches do not perform well for severe occlusions and extremely cluttered scene background. Finally, in prior work road features extracted from the image, e.g., edge and parallel boundary, are not sufficiently robust to be salient in all scenarios for road extraction.

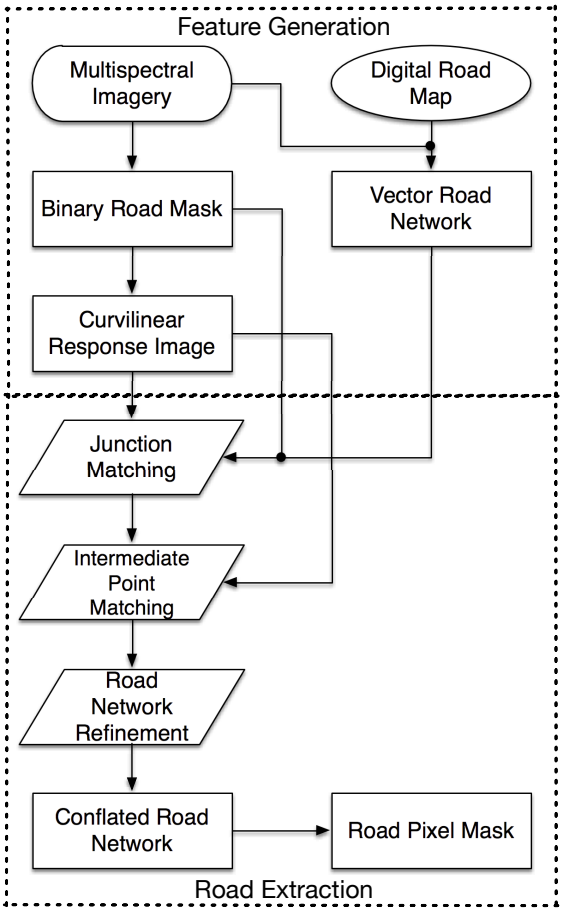


Figure 1.4: *System workflow layout for map conflation and image-based road extraction. The upper portion of this diagram describes data input and the stage for feature generation; the lower portion describes the stage for road extraction and map conflation and system output.*

1.3 Thesis Organization

Fig. 1.4 gives an overview of the whole system workflow to accomplish the objectives of this dissertation. A multispectral image is used to generate a corresponding binary road mask using one of the feature extraction techniques. From the binary road mask, a maximum projected curvilinear response image is created. Road vectors are concurrently generated from a digital road map. Based on the templates derived from the vector road network, junction matching is applied

to both the binary road mask and the curvilinear response image. The junction corrected road network is then combined together with the curvilinear response image to perform intermediate point matching. The conflated road network is finally converted to a road pixel mask.

The remainder of the dissertation is organized as follows. Chapter 2 presents a novel spectral similarity measure that is later used for road feature extraction; Chapter 3 introduces several techniques that are applied in generating unique and useful road features; Chapter 4 details our proposed approach for map-to-imagery conflation and image-based road extraction. Experimental results of applying our algorithm and quantitative evaluation are also provided; finally, the dissertation concludes in Chapter 5 with a summary and suggestions for future work.

Chapter 2

A Novel Spectral Similarity Measure for Urban Environments

The very first step of road extraction system is to create a good binary road mask, which has been highlighted in Fig. 2.1. Hence the quality of the mask is of significant importance to following processing steps, as well as the final outcome. There are plethora of ways to generate a binary road mask, but the most universal and robust way is the manual collection of a few representative road pixel samples from the image. Even for this basic method, many options are existent as to select an appropriate spectral similarity measure for comparison between reference pixel spectra and target image pixel spectra. To account for the unique characteristics of urban image objects, e.g., roads and building rooftops, a novel spectral similarity measure is proposed in this chapter to serve as an alternative measure that is best suitable for urban environments.

2.1 Introduction

In context of remote sensing, imaging spectroscopy reveals the unique property of the imaged target, which can be exploited to identify all objects with similar spectral characteristics. Spectral imaging from remote platforms can provide

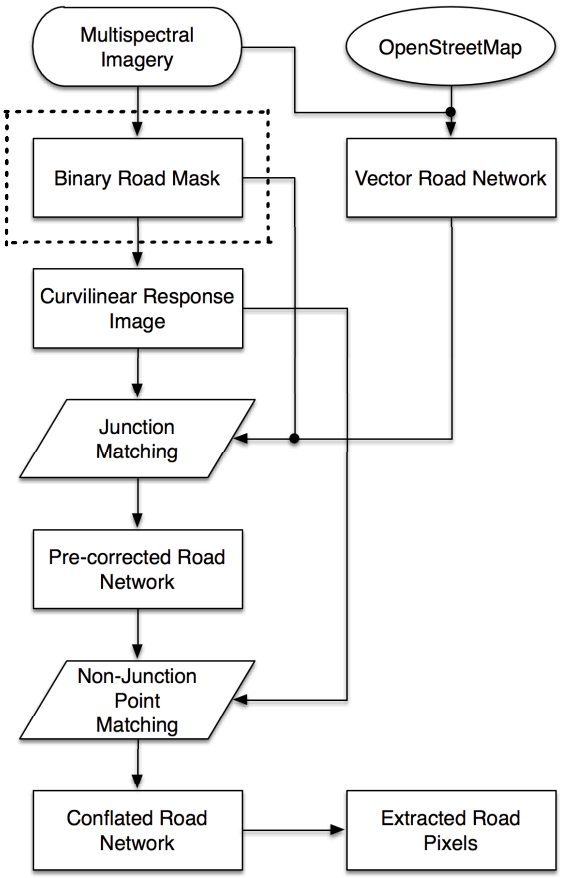


Figure 2.1: *System workflow for map conflation and image road extraction.*

periodic, large-scale, sustainable, and efficient surveillance of the earth surface and is crucial to target detection and land mapping efforts. The mapping process is often based upon the quantitative evaluation of the spectral similarity or dissimilarity of a target and a reference object in terms of a single pre-defined criterion, e.g., spectral angle mapper (SAM) [31], spectral information divergence (SID) [32, 33], spectral correlation mapper (SCM) [34, 35], squared Euclidean distance (ED), and squared Mahalanobis distance (MD) [36]. The reference spectral data can either be collected from field measurements, obtained from a spectral library, or extracted directly from the image. These measures are important tools in remote sensing imagery analysis and generic algorithm development. There are several prior publications describing applications and comparisons of spectral similarity measures [37, 38]. These works have demonstrated that spectral similarity measure can be very useful, but is often scene dependent.

Recent advances in sensing technology have led to improved spatial and spectral capabilities for remote sensing systems and enabled new detailed applications in urban areas. Small objects, e.g., architectural details, vehicles, signage, and even pedestrians are observable in urban settings. Yet the advanced resolving capability gives rise to the problem that excessive details, i.e., many unique man-made materials and local spectral variation caused by the topographical illumination effect [31, 39] (shadows and shading due to varied surface normals), are apparent in a complex scene. The non-uniform illumination issue can be solved using brightness normalization, but this only works on reflectance data. For example, to most effectively apply SAM, SID, and possibly SCM, the data should be reduced to "apparent reflectance", with all dark current and path radiance biases removed. [31] However, surface reflectance is not easily obtainable for high-resolution commercial satellite sensors. The data may only be available as digital number (DN) or uncalibrated spectral radiance data. The conversion of DN to reflectance requires ground truth data collected at the moment when the sensor platform flies overhead, or alternatively, some algorithms, e.g., Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) [40], QUick Atmospheric Correction (QUAC) [41], and Landsat Ecosystem Disturbance Adaptive Processing

System (LEDAPS) [42] can compensate for atmospheric effects without in situ data, but the fidelity cannot be guaranteed and the added complexity hinders their application.

Another issue in the analysis of high-resolution urban scenes is that many man-made materials have very similar spectral signatures. For example, gray asphalt roads and parking lots are often confused with bright roofing materials by spectral similarity measures that operate only on spectral direction, whereas magnitude (brightness) can be used to separate the materials. A measure such as `MD` is more appropriate in this case but the effort to find sufficient representative training data for each material in complex scenes becomes overwhelming.

One way to overcome the inherent limitations of single similarity measures is to combine them into hybrid measures. Efforts have been devoted to seek more accurate and robust spectral similarity measures by taking advantages of multiple measures. In [43], a comprehensive similarity index - spectral similarity value (SSV) - was generated by nonlinear combination of spectral magnitude, ED, and profile similarity, SCM. In [44], Du et al. proposed to combine variants of SAM (tangent or sine of SAM) and SID together by multiplication to form a new measure which yields a significant improvement over either SAM or SID. The product of SAM and SID considerably enhance the spectral discriminability, because it makes two similar spectra even similar and two dissimilar spectra more distinct. [45] also proposed to use a hybrid measure combined by SAM and cosine of the angle of SCM (SCA) for discrimination among *Vigna* species. [46] presented a new approach to change vector analysis based on both spectral direction and spectral magnitude. SAM and SCM are combined to compute spectral direction of change, while ED and MD are used to compute the magnitude component of spectral similarity. [15] developed a new measure by imposing boolean operations on SAM and ED to exploit spectral information for road network detection. [47] proposed a learnable hyperspectral measure by using multiple spectral measures, including SAM, SID, SCM, SSV, ED, and MD, as similarity patterns to train a classifier, which acts as an adaptive similarity threshold instead of a static one. However, hybrid measures may inherit the weakness of each combined measure,

determining how to combine measurements with different physical meanings (e.g., degree vs. distance) is a challenge, and training of optimal relative weight for each measure component may be burdensome. Further readings can be found in the literature [39, 48, 49].

Without reverting to a hybrid measure, one promising way to include both spectral direction and magnitude is to use the squared Mahalanobis distance, however, as previously mentioned, the training requirements to build the necessary covariance matrix are excessive for applications in complex environments. In this chapter we describe a novel spectral similarity measure that is built similar to `MD` but rather than collecting training data used to develop the covariance matrix, a user finds a few representative target pixels and then tunes an adjustable shape parameter for the target hyper-ellipsoid in spectral space. We also describe a simple user directed dark object subtraction method for calibration to apparent radiance to be applied to the data prior to the application of a spectral similarity measure. We then test our novel measure against a set of other single spectral similarity measures using receiver operator characteristic (ROC) curves to assess their performances. The remainder of the chapter is organized as follows: in Section 2.2, the overview of the common spectral similarity measures is given, followed by presenting a new radiometric calibration method and a new spectral similarity measure in Section 2.3; in Section 2.4 experimental results are provided for comparison; and finally the chapter concludes in Section 2.5.

2.2 Review of Single Spectral Similarity Measures

In this section we review the mathematical definitions and geometrical descriptions of the common single similarity measures presented in the previous section. This overview provides the context and mathematical basis that leads to the development of our new similarity measure. The pixel spectrum is mathematically represented as a vector in N-dimensional spectral space, where N is the number of spectral bands.

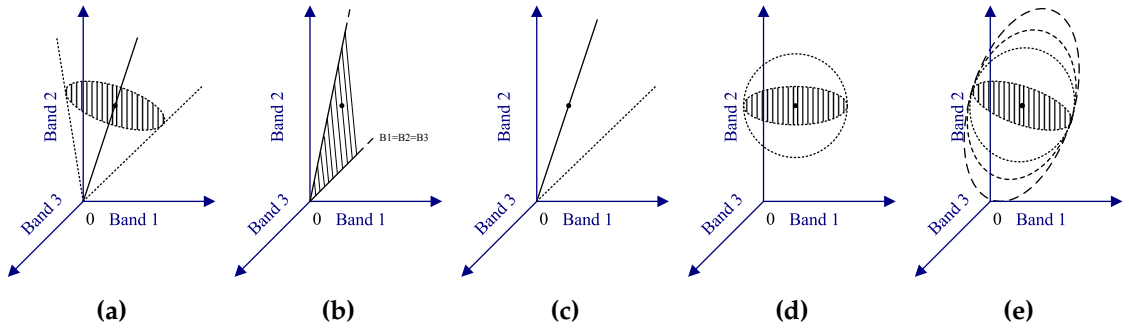


Figure 2.2: *Graphical isovalue surfaces of five spectral similarity measures. The black dots indicate the positions of arbitrarily chosen reference pixel points. The solid lines or the surfaces encompassed by solid lines represent the sets of points most similar to the reference in terms of similarity measures, while the outer surfaces encompassed by the dotted lines denote the sets of isovalue regions. Here "most similar" means having extreme spectral similarity scores. (a) SAM. (b) SCM. (c) SID. (d) ED. (e) MD with three different isosurfaces, which is equivalent to the proposed measure.*

2.2.1 Single Spectral Similarity Measures

Since it is impossible to visualize data with more than three dimensions, the isosurfaces of the similarity measures are depicted in a 3-D spectral space in Fig. 2.2. Note that the isovalue contours are not completely depicted in Figs.2.2b and 2.2c due to the undetermined locus of SCM and SID. Note that the surfaces of the isovalue contours must be bounded in the first quadrant due to the positivity of realistic physical quantities. $\mathbf{x}$ and $\mathbf{y}$ are defined as two vectors of random variables such that $\mathbf{x} = (x_1, x_2, \cdots, x_N)^{\mathrm{T}}$, $\mathbf{y} = (y_1, y_2, \cdots, y_N)^{\mathrm{T}}$, with the same dimensionality of N, where T denotes the operation of transposition.

2.2.1.1 Spectral Angle Mapper

SAM is an extensively used metric to evaluate the spectral similarity of a pair of two pixel spectra and is defined as

$$\mathrm{SAM}\left(\mathbf{x}, \mathbf{y}\right) = \cos^{-1}\left(\frac{\mathbf{x}^{\mathrm{T}}\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}\right). \tag{2.1}$$

The score given by SAM is normally measured in radians. The loca of pixel points with zero SAM value are depicted in Fig. 2.2a as a solid line running from the origin. The reference point also lies exactly on this straight line. An iso-value surface is overlaid as the dotted lines/contours and forms an upside-down cone with the origin as the vertex, but expanding to infinity. These characteristics show that SAM measures only the angular difference of spectral direction rather than the magnitude or brightness, because it is invariant to the scaling of spectral magnitude.

2.2.1.2 Spectral Correlation Mapper

SCM is defined as Pearson's correlation coefficient based on population statistics:

$$\mathrm{SCM}\left(\mathbf{x}, \mathbf{y}\right) = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}}, \tag{2.2}$$

where $\bar{x}$ and $\bar{y}$ are mean values of the components of two comparative vectors. SCM takes into account brightness differences and shape differences between spectra [37]. The correlation can either be positive or negative, but its maximum cannot exceed 1, which means fully (linearly) correlated. SCM is only sensitive to a linear relationship between two random variables. Pixel points with SCM values of 1 lie on a subset of a plane determined by the reference point and a space line $B1 = B2 = B3$ as highlighted by a series of parallel segments in Fig. 2.2b. SCM is invariant both to the spectral direction and translation of spectral magnitude.

2.2.1.3 Spectral Information Divergence

SID is defined as symmetrized discrete-form Kullback-Leibler (K-L) divergence by adding together two conjugate K-L terms:

$$\mathrm{SID}\left(\mathbf{x},\mathbf{y}\right)=\sum_{i=1}^{N}P\left(x_i\right)\log\frac{P\left(x_i\right)}{P\left(y_i\right)}+\sum_{i=1}^{N}P\left(y_i\right)\log\frac{P\left(y_i\right)}{P\left(x_i\right)}, \tag{2.3}$$

where the probability mass functions are defined as the normalized pixel spectra such that

$$P\left(x_i\right)=\frac{x_i}{\sum_{j=1}^{N}x_j},\ P\left(y_i\right)=\frac{y_i}{\sum_{j=1}^{N}y_j},\ i=1,2,\cdots,N. \tag{2.4}$$

If two random variables $\mathbf{x}$ and $\mathbf{y}$ have distinct probability distributions, SID will tend to be large. SID is applied to measure the spectral similarity between two pixel spectra $\mathbf{x}$ and $\mathbf{y}$ in a stochastic manner and is essentially the divergence based on the discrete probability of each band per pixel. Owning to the normalization, the pixel points with the zero SID value form a ray starting from the origin and passing through the reference point (see Fig. 2.2c), which is analogous to SAM. Given the difficulty of depicting an isosurface of SID, only one easily derived isovalue curve (dotted line in Fig. 2.2c) is drawn instead for illustration. SID is also invariant with the scaling of spectral magnitude.

2.2.1.4 Squared Euclidean Distance

ED is computed as the $L2$-norm difference of magnitudes of two spectra:

$$\mathrm{ED}\left(\mathbf{x},\mathbf{y}\right)=\left(\mathbf{x}-\mathbf{y}\right)^{\mathrm{T}}\left(\mathbf{x}-\mathbf{y}\right). \tag{2.5}$$

ED is an isotropic metric and does not favor any specific spectral direction. As shown in Fig. 2.2d, its isosurface is spherical with the centroid as the reference

point and the radius as the square root of its score.

2.2.1.5 Squared Mahalanobis Distance

MD is a class separability measure under the equal covariance and multivariate Gaussian assumption and is defined as

$$\mathrm{MD}\left(\mathbf{x},\mathbf{y}\right)=\left(\mathbf{x}-\mathbf{y}\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathbf{x}-\mathbf{y}\right), \tag{2.6}$$

where the covariance matrix Σ is typically learned from labeled data [50, 51, 52].

There is only a single point, which is the reference, with the similarity score of zero as illustrated in Fig. 2.2e. The isosurface of MD is an elongated (or equivalently squashed) ellipsoid along the direction of the reference vector. The length of principal axis is larger than that of any other minor axis with equal length.

2.2.2 Remarks

SAM and SID should only be used for evaluation of spectral similarity in terms of reflectance and they are insensitive to the lengths of spectral vectors. To the contrary, ED makes no distinction on spectral direction and merely records the relative $L2$ distance depending only on the spectral magnitude. MD, however, considers both spectral direction and spectral magnitude and emphasize according to covariance information. But MD is a class separability measure and cannot be used on spectral comparison of two single vectors. Although the statistical assumption of SCM is sound, the interpretation of its physical basis is not easy. It is difficult to explain why the pixel points most spectrally similar (fully correlated) to the reference point are located on an asymmetric subspace with respect to the reference vector (2-D plane in 3-D space). Further, its invariance with translation of magnitude is not physically reasonable. SID has a similar problem that its isosurface shows no symmetry, while SAM, ED, and MD are at least symmetric with respect to the line passing through both origin and reference point.

ED, SID, and MD are actually instances of Bregman divergence [53] with different convex functions. Bregman divergence is similar to a metric, but does not necessarily satisfy the triangle inequality nor symmetry. $F : \Delta \rightarrow \mathrm{R}$ is a continuously-differentiable real-valued and strictly convex function defined on a closed convex set Δ. Bregman divergence associated with F for points $\mathbf{x}, \mathbf{y} \in \Delta$ is

$$B_F(\mathbf{x} \| \mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - (\mathbf{x} - \mathbf{y})^{\mathrm{T}} \nabla F(\mathbf{y}), \tag{2.7}$$

where ∇ is the gradient operator. Eq. 2.7 can be thought of as the divergence between the value of F at point $\mathbf{x}$ and the value of the first order Taylor expansion of F around point $\mathbf{y}$ evaluated at point $\mathbf{x}$. Bregman divergence is a generalized measure that can be readily extended to some familiar forms.

ED can be easily generated by substituting the convex function $F(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x}$ into Eq. 2.7.

If we have the convex function with discrete form

$$F(\mathbf{x}) = \sum_{i=1}^{N} p(x_i) \log p(x_i), \tag{2.8}$$

where $p(\mathbf{x})$ is the probability mass function that suffices the probabilistic constraints $p(x_i) > 0, \Sigma_i p(x_i) = 1, i = 1, 2, \cdots, N$. By substituting $F(\mathbf{x})$ into Eq. 2.7, the corresponding Bregman divergence is

$$B_F(\mathbf{x} \| \mathbf{y}) = \sum_{i=1}^{N} p(x_i) \log \frac{p(x_i)}{p(y_i)}. \tag{2.9}$$

It actually reduces to discrete-form K-L divergence, which is also known as the relative entropy of $\mathbf{y}$ with respect to $\mathbf{x}$. K-L divergence basically measures how much information $\mathbf{x}$ has given the known probability distribution of $\mathbf{y}$. Intuitively speaking, if two random variables $\mathbf{x}$ and $\mathbf{y}$ have distinct probability distributions, K-L divergence will tend to be large. Here the probability mass functions are defined as the normalized pixel spectra such that

$$p(x_i) = \frac{x_i}{\sum_{j=1}^{N} x_j},\ p(y_i) = \frac{y_i}{\sum_{j=1}^{N} y_j},\ i = 1, 2, \cdots, N. \tag{2.10}$$

Probabilistic constraints are guaranteed, although they are not real probability measures. SID is constructed by adding together two conjugate K-L terms:

$$\mathrm{SID}(\mathbf{x}, \mathbf{y}) = B_F(\mathbf{x} \| \mathbf{y}) + B_F(\mathbf{y} \| \mathbf{x}). \tag{2.11}$$

In another case, if the convex function $F(\mathbf{x})$ is given as a general quadratic form $F(\mathbf{x}) = \mathbf{x}^{\mathrm{T}} Q \mathbf{x}$, the corresponding Bregman divergence is then given as

$$B_{F,Q}(\mathbf{x} \| \mathbf{y}) = (\mathbf{x} - \mathbf{y})^{\mathrm{T}} Q (\mathbf{x} - \mathbf{y}), \tag{2.12}$$

which is a generalization of the preceding squared Euclidean distance. If Q can be expressed as an inverse covariance matrix, i.e., Σ^{-1}, the Bregman divergence amounts to MD that is defined as

$$\mathrm{MD}(\mathbf{x}, \mathbf{y}) = B_{F,\Sigma^{-1}}(\mathbf{x} \| \mathbf{y}) = (\mathbf{x} - \mathbf{y})^{\mathrm{T}} \Sigma^{-1} (\mathbf{x} - \mathbf{y}). \tag{2.13}$$

Surprisingly, MD is essentially a special case of K-L divergence under Gaussian assumption. Suppose that the probability density function $p(\mathbf{x})$ is Gaussian N $(\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$, and $p(\mathbf{y})$ is another Gaussian N $(\mu_{\mathbf{y}}, \Sigma_{\mathbf{y}})$. The computation of K-L divergence is simplified as [54]

$$d_{\mathbf{x},\mathbf{y}} = \frac{1}{2}\mathrm{trace}\left(\Sigma_{\mathbf{x}}^{-1}\Sigma_{\mathbf{y}} + \Sigma_{\mathbf{y}}^{-1}\Sigma_{\mathbf{x}} - 2I\right) + \frac{1}{2}(\mu_{\mathbf{x}} - \mu_{\mathbf{y}})^{\mathrm{T}}\left(\Sigma_{\mathbf{x}}^{-1} + \Sigma_{\mathbf{y}}^{-1}\right)(\mu_{\mathbf{x}} - \mu_{\mathbf{y}}). \tag{2.14}$$

If the covariance matrices of the two Gaussian distributions are identical, i.e., $\Sigma_{\mathbf{x}} = \Sigma_{\mathbf{y}} = \Sigma$, then Eq. 2.14 is further simplified to

$$d_{\mathbf{x},\mathbf{y}} = (\mu_{\mathbf{x}} - \mu_{\mathbf{y}})^{\mathrm{T}} \Sigma^{-1} (\mu_{\mathbf{x}} - \mu_{\mathbf{y}}) \doteq \mathrm{MD}(\mu_{\mathbf{x}}, \mu_{\mathbf{y}}), \tag{2.15}$$

which is the squared Mahalanobis distance between the corresponding mean vec-

tors. If only the single pixel vector case is considered, i.e., $\mu_{\mathbf{x}} = \mathbf{x}$ and $\mu_{\mathbf{y}} = \mathbf{y}$, the divergence becomes the general squared Mahalanobis distance as shown in Eq. 2.13.

In summary, MD is simply the ED between two points transformed by a matrix dependent on Σ and MD is also a special case of SID under continuous Gaussian assumption of identical covariances and means as single pixel vectors. [54] Given that SID only sets probability measures as "normalized" spectral vectors, it seems that MD gives a richer representation of the underlying statistical distribution of the reference datum point.

Mathematical limitations of single spectral similarity measures lead to practical constraints: ED cannot handle spatially varying illumination issue; SAM, SID, and possibly SCM, can fix this but also require reflectance data. They also are weak in identifying objects with similar spectral profiles; MD cannot be used without a large training set. These limitations restrain the application of existing similarity measures to high-resolution urban scenes.

2.3 Proposed Scheme

Since reflectance data from high-resolution commercial satellite sensors are not routinely available, in Section 2.3.1, we present a simple radiometric calibration approach that can effectively preserve data collinearity to enable correction for the illumination effect. Then in Section 2.3.2, we describe a novel spectral similarity measure with feature similar to MD where both spectral direction and spectral magnitude are captured in a single measure.

2.3.1 Radiometric Conversion Method

In reflectance space, image data of the same type of material should distribute along a ray passing through the origin, and thus is the ideal situation for the application of similarity measures that leverage spectral direction. High-resolution

image data, however, are normally recorded as DN or radiance and include an atmospheric component. Further, high-fidelity radiometric calibration is not readily available for these images. To tackle the problem, a radiometric conversion approach, analogous to the well-known empirical line method (ELM), is proposed to perform atmospheric compensation and facilitate spectral analysis.

If a target lies on a slanted plane or its sky dome is obstructed by adjacent objects, the downwelled radiance onto the target will be reduced. Referring to Fig. 2.3, the fraction of the sky hemisphere above the target, also known as shape factor, is defined in accordance with geometry as

$$\mathscr{F} = 1 - \frac{1}{2}\sin\sigma_{\mathscr{F}}, \tag{2.16}$$

where $\sigma_{\mathscr{F}}$ is the target slope angle. Scene geometry relation is established with regard to $\sigma_{\mathscr{F}}$:

$$\sigma'\left(\mathscr{F}\right) = \sigma'_s - \sigma_{\mathscr{F}}, \tag{2.17}$$

where σ'_s is the solar zenith angle and σ' is the angle from the normal of the target to the sun.

The effective radiance at the sensor aperture, also known as top-of-atmosphere (ToA) radiance, is effectively approximated as [2]

$$L_{ToA}\left(\lambda\right) = \left[E'_s\left(\lambda\right)\cos\sigma'\left(\mathscr{F}\right)\tau_1\left(\lambda\right) + \mathscr{F}E_d\left(\lambda\right)\right]\frac{r\left(\lambda\right)}{\pi}\tau_2\left(\lambda\right) + L_u\left(\lambda\right), \tag{2.18}$$

where $E'_s\left(\lambda\right)$ is the exoatmospheric spectral irradiance onto a surface perpendicular to the incident beam, $\tau_1\left(\lambda\right)$ is the atmospheric transmission along the sun-target path, $E_d\left(\lambda\right)$ is the total downwelled spectral irradiance, $r\left(\lambda\right)$ is the bidirectional reflectance factor, $\tau_2\left(\lambda\right)$ is the atmospheric transmission along the target-sensor path, $L_u\left(\lambda\right)$ is the radiance from the sun scattered upward into the sensor's line of site along the sensor-target path. Since the physical parameters are all dependent on the wavelength, the explicit dependency of λ will be removed for brevity.

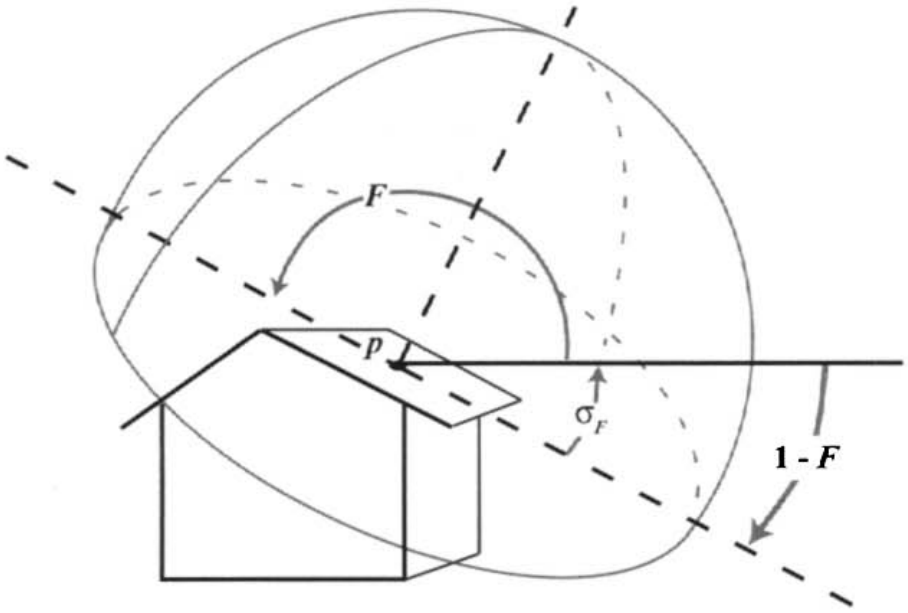


Figure 2.3: *Shape factor $\mathscr{F}$ of the exposed sky. The slope of roof reduces the amount of sky seen by the point p. (image adapted from [2])*

ELM [55, 56] is a linear regression of observed radiance values against known reflectance values (ground truth) according to

$$L_{ToA} = \left[E_s' \cos\sigma'\left(\mathscr{F}\right)\tau_1 + \mathscr{F}E_d\right]\tau_2\pi^{-1}\cdot r + L_u, \tag{2.19}$$

where the leading factor before r is the slope of the regression and L_u is the intercept. The image data can then be calibrated to surface reflectance based on this linear relationship. ELM assumes that ToA radiance is proportional to the varied reflectance. For our application, constant reflectance of a single object is desired, but the true reflectance value may remain unknown. We would like to relate radiance data to the shape factor $\mathscr{F}$, rather than the reflectance r, so as to achieve consistent spectral similarity measurements with various scene geometries, which result in non-uniform illumination. τ_1, τ_2, and L_u are assumed to be constant over the scene as long as the sensor does not have a very wide field of view and a very large area of coverage. In addition, we assume that the samples are approximately Lambertian so that the errors introduced by sensor viewing angle effects could be minimized. If this is not the case, solving for the reflectance value r requires estimating the shape of the bidirectional reflectance distribution function (BRDF). [57, 58]

If the relative amount of skylight to the total direct solar irradiance incident on the target is known, i.e., $E_d = \alpha E_s'\tau_1$, where α is a scaling factor for each band. Eq. 2.19 can be subsequently reduced:

$$\begin{aligned} L_{ToA}\left(\mathscr{F}\right) &= \left[\cos\sigma'\left(\mathscr{F}\right) + \mathscr{F}\alpha\right]E_s'\tau_1\tau_2\pi^{-1}\cdot r + L_u \\ &\equiv A\cdot\beta\left(\mathscr{F}\right) + B, \end{aligned} \tag{2.20}$$

where $\beta\left(\mathscr{F}\right) = \cos\sigma'\left(\mathscr{F}\right) + \mathscr{F}\alpha$, $A = E_s'\tau_1\tau_2\pi^{-1}r$, and $B = L_u$. Derived from Eq. 2.20, the ground-leaving radiance of the target is then expressed as the difference

between the ToA radiance and the upwelled radiance,

$$L_{gd} \equiv L_{ToA}(\mathscr{F}) - B = A \cdot \beta(\mathscr{F}). \tag{2.21}$$

Suppose that two targets, T_1 and T_2, consisting of the same material lie on a surface with varied surface normals, which leads to different shape factors $\mathscr{F}_1$ and $\mathscr{F}_2$, respectively. From Eq. 2.21, The ratio of ground radiance of two samples is then given:

$$\frac{L_{gd,T_1}}{L_{gd,T_2}} = \frac{L_{ToA}(\mathscr{F}_1) - B}{L_{ToA}(\mathscr{F}_2) - B} = \frac{\beta(\mathscr{F}_1)}{\beta(\mathscr{F}_2)} = \frac{\cos\sigma'(\mathscr{F}_1) + \mathscr{F}_1\alpha}{\cos\sigma'(\mathscr{F}_2) + \mathscr{F}_2\alpha}. \tag{2.22}$$

Note that α in Eq. 2.22 is implicitly dependent on λ. If we relax the constraint on band-dependency and the scaling factor no longer depends on wavelength λ, i.e.,

$$\alpha(\lambda) = \alpha, \tag{2.23}$$

Eq. 2.22 will yield a constant value. The ratio is constant across all spectral bands for any two arbitrarily given target pixels representing exactly the same material. Now ground-leaving radiance has properties similar to reflectance, and spectral similarity measures can be applied to image data with upwelled radiance removed, because the data collinearity is preserved.

There are several techniques for estimating the upwelled radiance. Since we assume that there are no ground truth data available, we implemented - dark object subtraction (DOS) - to correct for the atmospheric effect. DOS assumes the existence of a dark object with zero or small surface reflectance in the scene and a horizontally homogeneous atmosphere. [59] The minimum ToA radiance value in the histogram from the entire scene is thus attributed to the atmospheric effect and is subtracted from all the pixels [60]. We simply seek to find one of the darkest pixels in the scene which is defined as the one without direct solar illumination and also receiving negligible sky light illumination. As a result, deeply shadowed pixels surrounded by high-rise buildings and/or tall tress are potentially feasible candidates for an urban dark object and can be easily identified

by the user. Hence, ToA radiance of the dark pixel is contributed solely by the upwelled radiance and Eq. 2.18 reduces to a much simpler form:

$$L_{ToA,dark} = L_u = B. \tag{2.24}$$

The cumulatively upwelled radiance is assumed to be uniform across the whole image scene. It can be seen that ToA radiance data of two targets T_1 and T_2 are not collinear with a line passing through the origin due to the offset caused by the upwelled radiance. Only after L_u is subtracted from the ToA radiance can the data collinearity be found. This type of data could also be possibly used to match with a spectral library that is mainly documented as calibrated data.

2.3.2 Proposed Spectral Similarity Measure

We previously discussed how MD has the advantage of capturing more spectral features, but the disadvantage of requiring the collection of a large training data set to construct a nonsingular covariance matrix. It is challenging to find adequate representative spectral training pixels in a complex urban scene, where there are too many variants within a single class. To exploit spectral features as much as possible and lessen the dependence on training, a new spectral similarity measure is introduced based upon MD but with a user adjustable isovalue hypersurface. It incorporates spectral direction and spectral magnitude naturally, because it does not require any kind of combination. In addition, the proposed approach offers more freedom in designing a distance isosurface in spectral space, which facilitates the manipulation of the comparative tendency towards either spectral direction or spectral magnitude.

Referring to Eq. 2.6, covariance matrix Σ is always symmetric and thus can be diagonalized as $\Sigma = \Phi\Lambda\Phi^{\mathrm{T}}$, where $N \times N$ orthonormal matrix Φ has as its columns the corresponding eigenvectors of Σ, and Λ is a $N \times N$ diagonal matrix whose diagonal elements are variances of the corresponding bands. Its first element is intentionally made w^2 times of the rest such that

$$\Lambda = \begin{bmatrix} w^2 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{2.25}$$

where $w \geqslant 1$ and $w \in \mathbb{R}$, thus it is assured that the first element w^2, corresponding to the reference vector $\mathbf{x}$ in Eg. 2.6, is reasonably larger than the other elements. It is implicitly assumed no preference for spectral direction other than the direction of the reference vector. By altering w, we can achieve different shapes and orientations of the isovalue hyper-curves in spectral space. The preference for the extent of spectral "purity" is literally controlled by the selection of w. Larger w means spectral direction dominates more over magnitude, which visually makes the hyper-ellipsoid "flatter" and the measure more sensitive to the change of spectral direction, and vice versa. The customized covariance matrix Σ can then be easily constructed because only the subspace spanned by the axial vectors normal to the reference matters, see

$$\Sigma = I + \left(w^2 - 1\right) \mathbf{x}\mathbf{x}^{\mathrm{T}} / \left\|\mathbf{x}\right\|^2, \tag{2.26}$$

where I is a $N \times N$ identity matrix. The construction only depends on the reference vector $\mathbf{x}$ and variable multiple w. Given Λ is of full rank, Σ is always invertible.

According to Eq. 2.26, the proposed measure is named anisotropy-tunable distance (ATD), because it uses a user-defined covariance matrix Σ, rather than one generated from training data in a traditional way. Since only linear operations are involved, the generation of a covariance matrix is not computationally expensive. The hyper-ellipsoidal isosurface of ATD is elongated along the principal axis, i.e., the direction of $\mathbf{x}$, which indicates its sensitivity to the change of spectral direction. The shape of the isosurface can be tuned by adjusting the first eigenvalue of Σ, upon which the extent of sensitivity to varying spectral angles is strengthened or weakened. This configuration results in a series of hyper-ellipsoidal isosur-

faces with respect to varied w (see Fig. 2.2e).

It is also straightforward to see that ATD is equivalent to ED when $w = 1$. Unlike the existing measures, choosing $\mathbf{x}$ or $\mathbf{y}$ as the reference will result in different measure scores for ATD. In other words, ATD is not a metric because it is not symmetric. But this is not a concern, since for most applications we compare pixel vectors with only one reference vector, which can be derived from the image manually or automatically. For example, one application of spectral similarity measures is active contour based image segmentation, [61] which uses mean intensity within a region as the spectral reference to be compared with pixel intensities within this region. The extraction of reference is image-based and fully automatic. In the following, we will use ATD(ω) to denote the similarity score of ATD when parameter is ω.

2.3.3 Summary

The proposed scheme involves a novel distance measure for evaluating spectral similarity in high-resolution urban scenes and an integrated radiometric calibration approach. The new measure is generated under the Gaussian assumption of data distribution and uses a single reference vector as the mean. The shape of isosurface is adjusted by changing the first eigenvalue of the covariance matrix. By doing this we can tune its inclination towards spectral direction or magnitude and make it adaptive to complex urban scenes. In addition, the radiance calibration is greatly simplified because no reflectance data are needed. Though no reflectance data are used, the collinearity of radiance data has been effectively preserved, which benefits any similarity measure which leverages spectral direction.

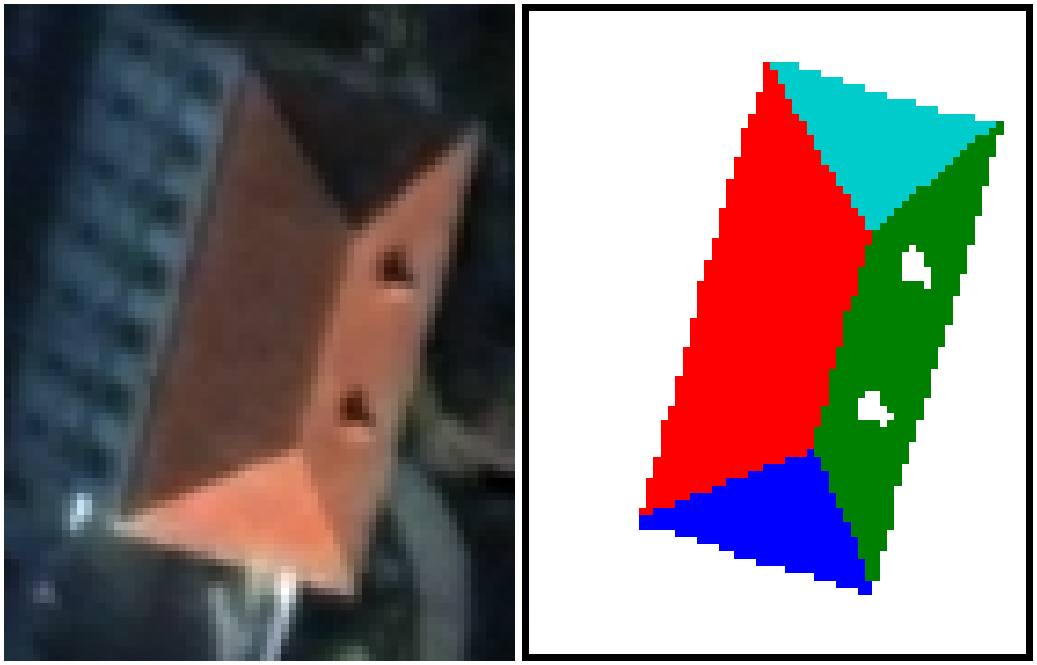
This approach may fail if the upwelled radiance data are spatially varying, the scene objects are far from Lambertian, or the scaling factor α changes dramatically from band to band. If the Gaussian assumption of data distribution about the reference no longer holds (though this is rare in real world), ATD is inappropriate for similarity assessment and other existing measures should be adopted instead.

2.4 Results & Discussion

Images used in the experiments are from two commercial high-resolution multispectral satellite sensors, GeoEye-1 (4 bands) and WorldView-2 (8 bands). They are pan-sharpened to 0.5-meter resolution, covering the spectral range from visible to near IR bands. In this section, we show the results of different spectral similarity measures applied on the imagery from both sensors. Based on the linear relationship, original DN data has been converted to ToA radiance for both sensors according to [62].

2.4.1 Data Collinearity Verification

The proposed scheme is first applied to a pan-sharpened WorldView-2 image. In order to test the effectiveness of DOS in preserving data collinearity in spectral space, an image portion (484×484) containing a typical hip roof with four sloped orientations is cropped from the original and is shown as natural color in Fig. 2.4a. Due to the slanted angles, sloped faces of the red roof receive different solar illuminations, resulting in distinct shading effects. Here it is assumed that this whole rooftop consists of a single material, ignoring small structures (e.g., two dark unresolvable objects on east face). Thus it is expected that data points of the rooftop should be collinear with respect to the origin in a multidimensional space. Four ground truth regions of interest (ROIs) are selected such that each region represents a unique illuminating condition, which are illustrated in Fig. 2.4b. The image data in RGB color space within these ROIs before and after the DOS processing are depicted in Fig. 2.5. Note that the original radiance is always larger than the adjusted radiance due to the positive contribution of the upwelled radiance. Mathematically, two separate straight lines passing through the origin are then fitted to two image data sets in a least-squared sense. As plotted in the same figure, the solid line denotes the fitted line for original data while the dashed line is fitted to post-DOS data. Brighter lines on the green-blue plane are the projections of the fitted lines. It is obvious to see that the adjusted radi-



(a) (b)

Figure 2.4: *(a) WorldView-2 pan-sharpened natural color image (12/10/2009) of downtown Rome, Italy showing a rooftop with four facial orientations, leading to different shading appearances. Colored pixels represent corresponding ground truth ROIs for data collinearity test. (image courtesy of DigitalGlobe) (b) shows labeled ground truth of a rooftop and each color represents one side of the roof.*

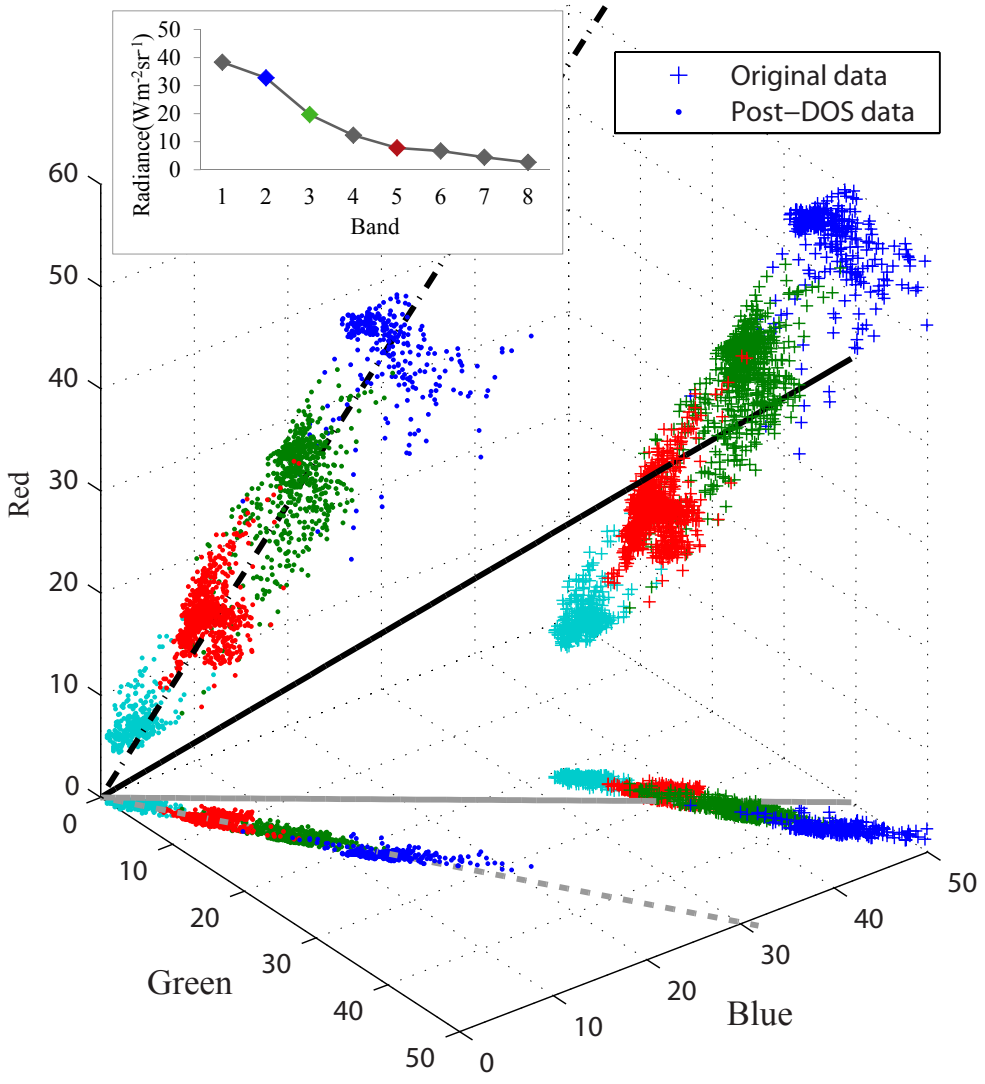


Figure 2.5: *Collinearity test of the image data before and after DOS. The colors of the data points correspond to ROIs in Fig. 2.4b. The marks '+' and '.' denote the original and adjusted radiance data respectively. The data points in RGB space are also projected onto the green-blue plane. R^2 is 0.69 for the post-DOS data and 0.20 for the original data. The dark pixel spectrum extracted from WorldView-2 image scene is shown as the inset.*

(a) (b)

Figure 2.6: *Detection of roof and road/parking lot. (a) GeoEye-1 pan-sharpened natural color image (06/04/2010) of a typical urban scene in southeast Rome, Italy, with a color composition of bands 3, 2 and 1. Yellow cross and green diamond indicate positions of selected reference pixel points for roofs and roads, respectively. (image courtesy of GeoEye) (b) labeled ground truth: red denotes rooftop and blue denotes roads/parking lots.*

ance data preserve much better collinearity than the original data, with R^2 values - 0.69 for post-DOS data and 0.20 for original data. The proposed radiometric conversion is justified because the collinearity of the radiance data with respect to the origin of a homogeneous material is greatly improved after DOS and the radiance data have been calibrated.

2.4.2 Spectral Similarity Measures Comparison

Referring to Fig. 2.6a, a 484×484 cropped portion of a GeoEye-1 pan-sharpened ToA radiance image is used to demonstrate the applicability of ATD and evaluate its performance compared with other spectral similarity measures. There are two reasons why this particular scene is chosen: this is a typical urban scene covering various types of objects such as buildings, paved roads, parking lots, trees, lawns,

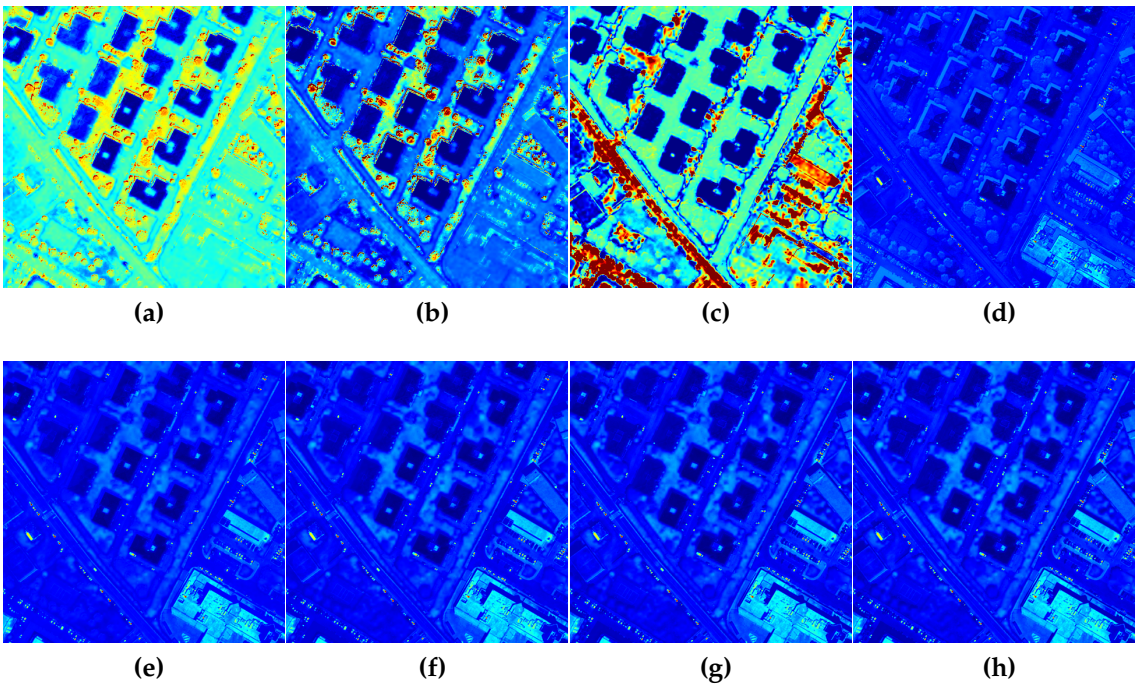


Figure 2.7: *For the GeoEye image, color-coded spectral similarity scores given by five spectral similarity measures with respect to the red pitched rooftops. Negative values of* `SCM` *are truncated to zeros and then subtracted from 1. (a) SAM. (b) SID. (c) 1-max(0,SCM). (d) ED. (e) ATD(3). (f) ATD(5). (g) ATD(7). (h) ATD(9).*

vehicles, etc.; on the other hand, within-class spectral variations for rooftops and roads/parking lots are visible. Two important applications of spectral similarity measures in urban scenes, detection of buildings and detection of roads/parking lots, will be examined for this scene.

About a dozen of building complexes with reddish rooftops are the objectives that we would like to separate from irrelevant objects and background clutter for purpose of, say, image segmentation. The reference vector is acquired by averaging over a few representative pixels on a typical red rooftop (see Fig. 2.6a). As shown in Fig. 2.7, warm color and cool color represent high or low spectral similarity scores respectively. (n.b., the scores of each spectral similarity are scaled if possible to help a better visualization.) All reddish roofs are expected to be rendered as dark blue, because low spectral similarity score indicates high spectral resemblance. In this case, this type of roofing is assumed to be uniform. All similarity measures yield somewhat anticipated outputs. Referring to Figs. 2.7a to 2.7c, SAM, SID, and SCM all produce strong spectral separation between targets and background with high local contrast, while ATD measures have weaker contrast between foreground and background. ED misses partial rooftops in some reddish-brown buildings due to the shading issue. As a whole, those measures that exclusively extract spectral direction perform best.

The GeoEye-1 image is used again to find spectrally similar pixels with respect to paved roads and connected parking lots. Referring to the green diamonds labeled in Fig. 2.6a, the reference vector is acquired by averaging over four representative pixels on paved roads. Here all road pixels are expected to appear as dark blue with other pixels having warmer colors. Referring to Figs. 2.8f and 2.8g, ATD(5) and ATD(7) both give reasonably good separation such that major roads, as well as residential roads, are assigned low scores while vegetation appears red or yellow, and other types of roofs are blue, but brighter. As w becomes larger and larger, see Figs. 2.8d ($w = 1$) to 2.8h ($w = 9$), ATD gives a weaker and weaker contrast between roads and rooftops (e.g., flat gray and white rooftops), but a better separation from vegetation (e.g., trees and lawns). Apparently paved roads and gray or white rooftops have similar spectral profiles, meaning after

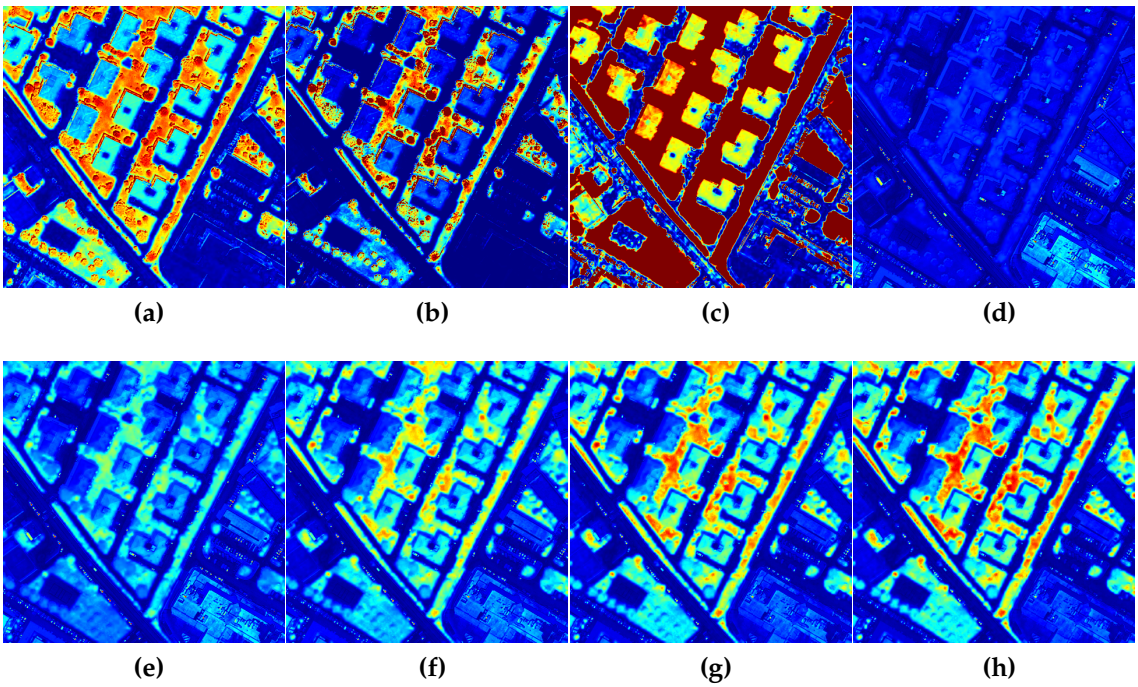


Figure 2.8: *For the GeoEye image, color-coded spectral similarity scores given by five spectral similarity measures with respect to the roads. Negative values of* `SCM` *are truncated to zeros and then subtracted from 1. (a) SAM. (b) SID. (c) 1-max(0,SCM). (d) ED. (e) ATD(3). (f) ATD(5). (g) ATD(7). (h) ATD(9).*

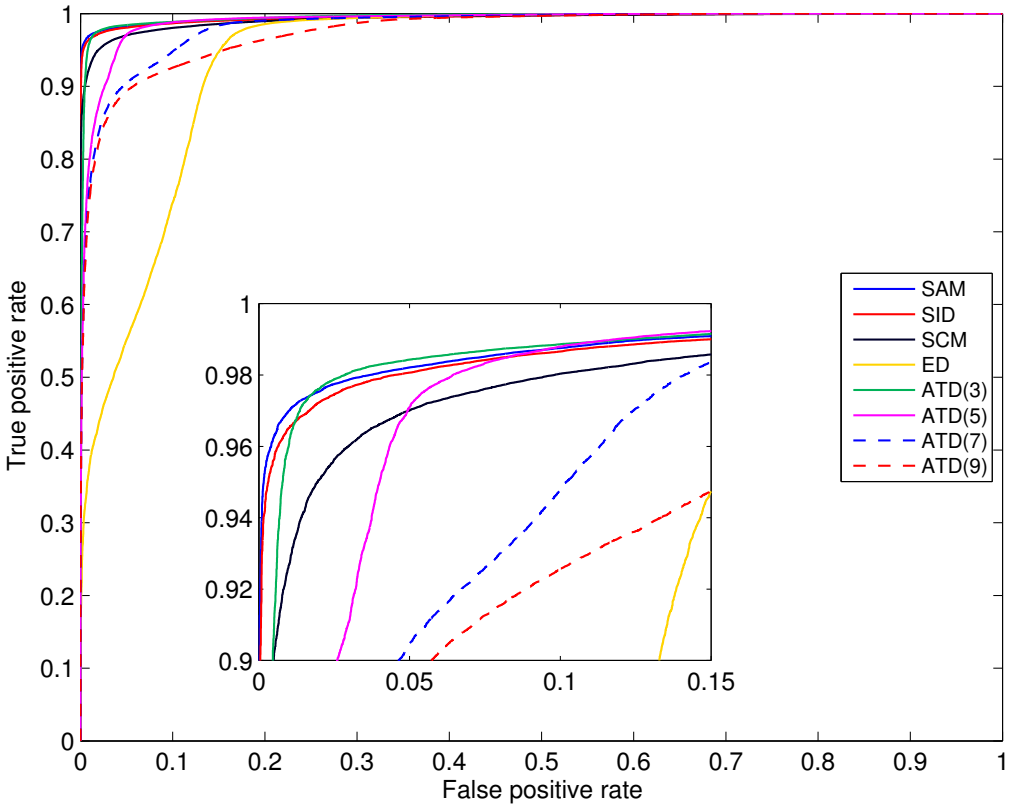


Figure 2.9: *ROC curves for roof detection.*

brightness normalization their spectra bear a strong resemblance. Consequently, the grayish roofing close to the right side of Fig. 2.6a is almost indiscernible from roads for ATD(9), see Fig. 2.8h, as well as SAM (Fig. 2.8a) and SID (Fig. 2.8b). SCM yields an unsatisfactory result since severe spectral confusions arise.

Next, given each spectral similarity measure, we threshold its score and generate a ROC curve against the ground truth data of either roof or road/parking lot (Fig. 2.6b). ROC curves of the detection of both urban objects for all spectral similarity measures are plotted in Figs. 2.9 and 2.10, respectively, and the corresponding area under curve (AUC) values are illustrated in Fig. 2.11. SAM, SID, and SCM all produce very high AUCs in roof detection, while their performances deteriorate in segmenting roads and parking lots, which verifies previous

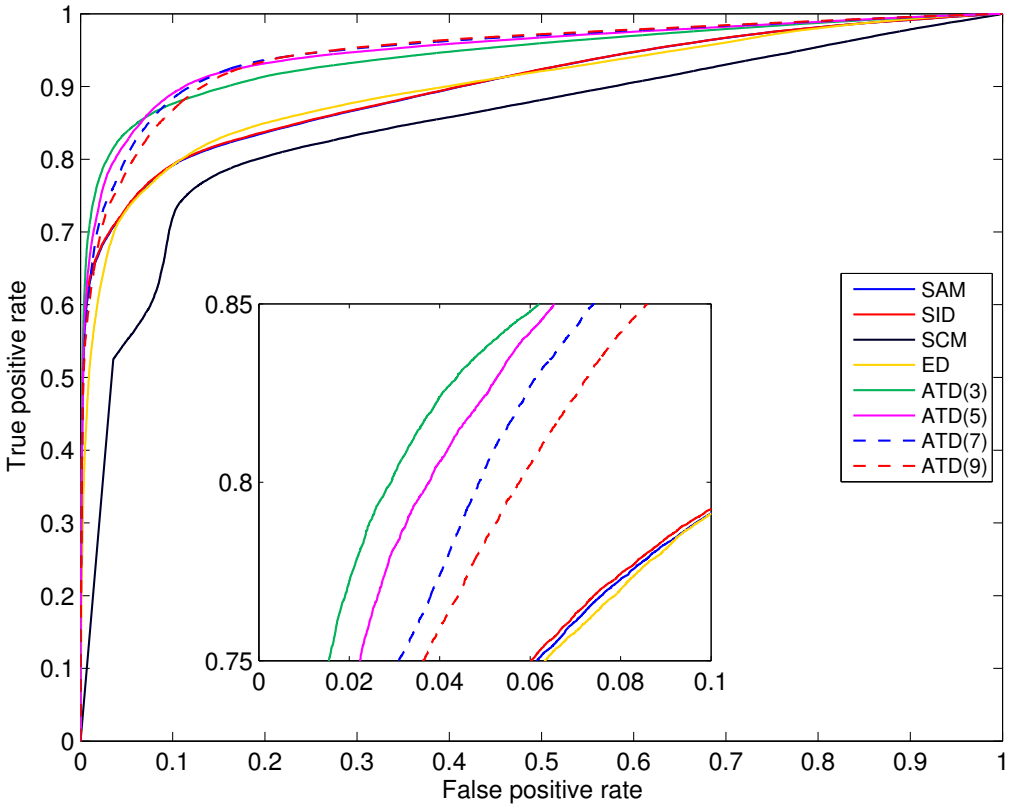


Figure 2.10: *ROC curves for road/parking lot detection.*

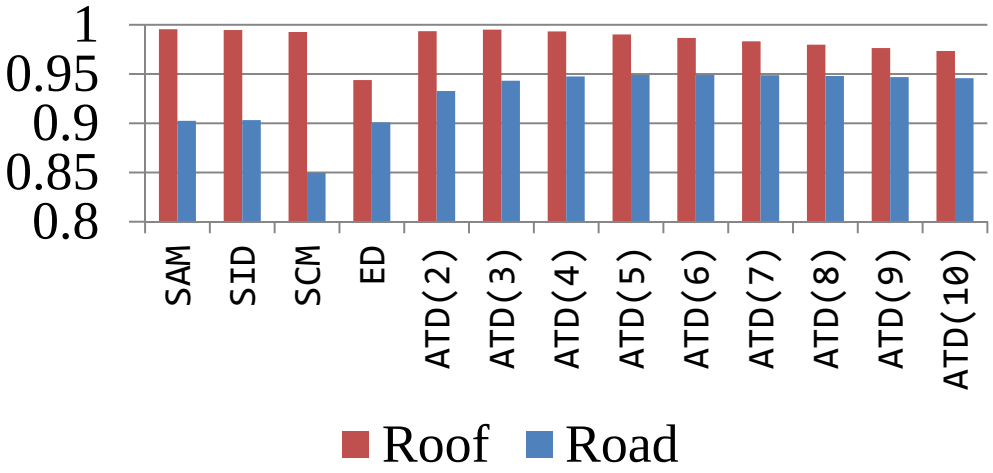


Figure 2.11: *AUC chart for roof and road detection.*

visual inspection. This is because the pixel intensity differences of red rooftops are caused by non-uniform illumination and can be compensated by brightness normalization. However the spectral heterogeneity of roads/parking lots is not caused by scene geometries, but rather distinct materials. Missing spectral magnitude knowledge weakens the separability of these measures. ATD(5) produces slightly lower AUC value in roof detection but higher AUC value in road detection because both spectral direction and magnitude are taken into account in a balanced manner. ATD has peak AUC values at $w = 3$ for roof detection and $w = 7$ for road detection, which can also be verified by ROC curves (see green curve in Fig. 2.9 and blue dashed curve in Fig. 2.10). There is always a balance to strike between competing goals like separating spectrally similar objects, either in terms of spectral direction or spectral magnitude. In this scene, $w = 5$ is optimal for detecting roofs and roads/parking lots simultaneously.

In conclusion, SAM and SID have similar performance in measuring similarity due to the common magnitude normalization process. SCM was not very powerful in consistently detecting urban structures. ED is one of the most commonly adopted spectral similarity measures. It is, however, unable to utilize the spectral direction information. ATD has a good balance of performance with sufficient flexibility and can be adapted to complex scenes at a low computational cost of a few trial-and-error tests for selecting the optimal parameter w. ATD yields satisfactory results such that non-uniform illumination of the same object is compensated and truly spectrally different objects are separated. Based on the ROC curve analysis, the optimal w value generally falls between 2 and 7.

Although the selection of w is not strictly scene-dependent, we envision that a user would do a few trials to determine the optimal value of w for a given scene and given application. However, an optimal solution of w may be difficult to find if the class data do not distribute along the direction of reference.

With regard to the computational time, all spectral similarity measures run fairly fast on the test image, (see Table 2.1). The tests were conducted using Matlab on a PC with a 2.67GHz CPU. Since the spectral similarity measurements work on a per-pixel basis, parallel computing can be easily deployed without

	SAM	SID	SCM	ED	ATD(7)
Fig. 2.7	29	93	44	8	42
Fig. 2.8	30	90	44	8	46

Table 2.1: *Computational time in millisecond (averaged over 10 runs).*

extra cost, simply by partitioning the image.

2.5 Conclusions

A modified dark object subtraction method using deeply shadowed areas common to urban scenes is employed to simplify radiometric conversion and reinforce the collinearity of image data with respect to the origin in multidimensional spectral space and thus account for the illumination effect. Even though it might not be optimal for detection of road class that typically does not have the same pavement material, DOS is necessary for rooftop that is processed with identical material. Commonly used single spectral similarity measures may not perform well at differentiating objects with similar spectral profiles, because they work on either spectral direction or magnitude, but not both. Though hybrid measures combining multiple measures are possible, they also inherit drawbacks from combined single measures. Our spectral similarity measure, ATD, is designed to behave similar to MD and to naturally and adaptively exploit relatively complete information. An examination of ROC curve results indicates that ATD is capable of evaluating spectral similarity in high-resolution urban scenes. The promising results demonstrate that ATD can provide a consistently reliable measurement of the spectral similarity in urban scenes. Future work involves assessment of its application to more complicated scenes and methods for integrating ATD with image segmentation techniques and other practical applications.

Chapter 3

Road Feature Generation

The road extraction system workflow is redrawn in Fig. 3.1 to highlight the feature extraction stage of our proposed map conflation and road extraction system. This chapter will cover those bounded building blocks in the upper portion of the workflow. Two independent data sources, namely a geo-referenced multispectral image and a digital road map, are fed into the proposed system along separate vertical paths, and the generated road features will be applied in a road extraction stage in Chapter 4.

An efficient and robust pan-sharpening algorithm, so called NNDiffuse, is introduced and applied in Section 3.1 as an optional preprocessing step. Along the left path in Fig. 3.1, the original or pan-sharpened multispectral image is used to generate a binary road mask, which reveals the unique image road features and will be discussed in Section 3.2. The binary road mask is then used to create a curvilinear structure response image and will be covered in Section 3.3.

Along the right path in Fig. 3.1, based on the image geographical projection metadata, the road map is registered to the image intrinsic coordinate system and becomes a vectorized road network. For the sake of convenience of image processing, the vector road map is further rasterized to match the regular image grid. This step will be discussed in Section 3.4. Finally, this chapter is summarized in Section 3.5.

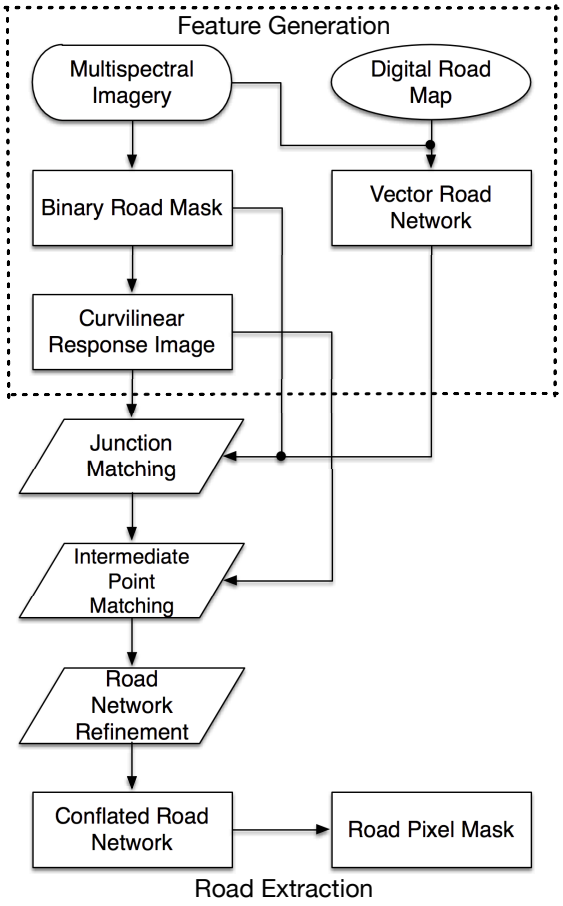


Figure 3.1: *System workflow for road feature extraction. Feature generation stage is focused on in this chapter. Related road features are extracted from the multispectral image and corresponding road map through the steps enclosed in the bounding box.*

3.1 Multispectral Image Pan-sharpening

Remotely sensed imagery is often delivered as an image pair, a high-resolution panchromatic image and a low-resolution multispectral image, because the on-board spectrometer must maintain a larger pixel size to be consistent with the SNR (signal-to-noise ratio) of the panchromatic sensor. The purpose of *pan-sharpening* is to produce a fused high-resolution multispectral image from the original image pair. For example, Landsat-8 OLI has eight 30-meter resolution multispectral bands and one 15-meter resolution panchromatic band. [63] Pan-sharpening will yield eight 15-meter resolution multispectral bands. This step is not explicitly listed in the workflow but sometimes it is desired to use a pan-sharpened multispectral image to resolve roadways better, rather than the original lower-resolution multispectral image.

Many pan-sharpening algorithms have been proposed over the past decades, but among them, the three most popular ones have different restrictions: the Gram-Schmidt method is efficient but cannot produce consistent high-quality fusion results; and the gMMSE method [64] and UNB sharpening method [65] are both proprietary. To provide an alternative solution, a novel nearest neighbor diffusion (NNDiffuse) based pan-sharpening algorithm which uses the per pixel spectrum and generates a resolution-enhanced multispectral image using a common linear mixture model is summarized below. The details of this algorithm are found in Sun et al. [3].

We use a typical image pair as an example to illustrate our algorithm. This image pair comprises of a full-size single-band panchromatic image and a quarter-size multi-band multispectral image. The difference factors $N_j(x, y)$ estimate the similarities of the pixel of interest (x, y) on a finer grid to its nine superpixels (u, v) on a coarser quadruple grid by comparing a summation of absolute difference. $N_j(x, y)$ from neighboring superpixels are acquired for each pixel from the

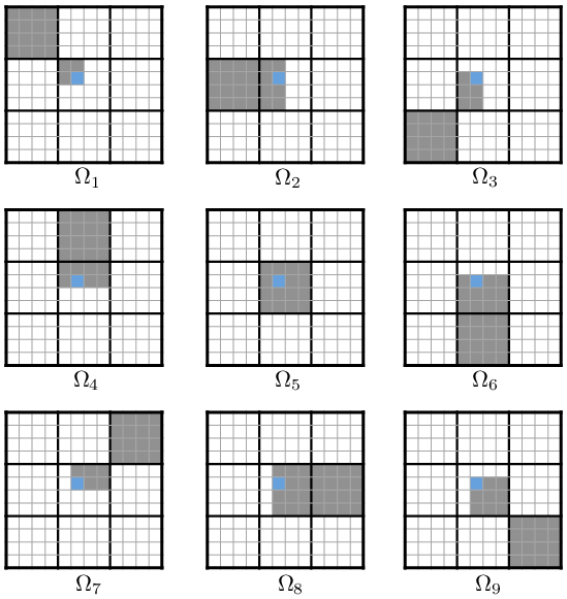


Figure 3.2: *Zoning illustration for fusion algorithm. Blue pixel represents the pixel of interest* (x, y) *and gray pixels are those pixels in the fusion zone* Ω_j*. Image adapted from [3].*

panchromatic image at the original resolution and are calculated from

$$N_j(x,y) = \sum_{(p,q)\in\Omega_j(x,y)} |P(x,y) - P(p,q)|, \qquad j = 1,2,...9, \tag{3.1}$$

where $P(\cdot,\cdot)$ represents the intensity value at any give pixel location in the panchromatic image, $\Omega_j(x,y)$ defines the diffusion region for each of the nine neighboring superpixels as shown in Fig. 3.2, and (p,q) represents any pixel within Ω_j. It is then possible to generate the fused spectrum $\mathbf{F}$ based on the idea of anisotropic diffusion [66] as

$$\mathbf{F}(x,y) = \frac{1}{k(x,y)} \sum_{j=1}^{9} \exp[-\frac{N_j(x,y)}{\sigma^2}] \times \exp[-\frac{\parallel (x,y) - (x_{u,v}, y_{u,v})|_{x,y,j} \parallel}{\sigma_s^2}] \mathbf{M}(u,v;x,y,j), \tag{3.2}$$

where $\mathbf{M}(u,v;x,y,j)$ is the pixel spectrum of each neighboring superpixel (u,v) corresponding to pixel (x,y) and j is in accordance with diffusion regions illus-

trated in Fig. 3.2. $(x_{u,v}, y_{u,v})$ is the center pixel location of each of the nine neighboring superpixels (u, v). σ and σ_s are intensity (range) and spatial smoothness factors that control the sensitivity of the diffusion. Eq. 3.2 relates diffusion factors to a multiplication of intensity similarity and spatial closeness. Inside the summation, $\exp[-\frac{N_j(x,y)}{\sigma^2}]$ gives a similarity measurement between (x, y) and its neighboring superpixels, while $\exp[-\frac{\|(x,y)-(x_{u,v},y_{u,v})|_{x,y,j}\|}{\sigma_s^2}]$ represents spatial proximity of (x, y) to the center of the neighboring superpixels. $k(x, y)$ is a normalization factor calculated as

$$k(x,y) = \frac{\sum_{j=1}^{9} \exp[-\frac{N_j(x,y)}{\sigma^2}] \times \exp[-\frac{\|(x,y)-(x_{u,v},y_{u,v})|_{x,y,j}\|}{\sigma_s^2}]\mathbf{M}(u,v;x,y,j) \times \mathbf{T}}{P(x,y)}, \tag{3.3}$$

where $\mathbf{T}$ is a spectral band photometric calibration vector. It is used to normalize the spectra values so as to make sure that panchromatic band is compatible with the linear combination of multispectral bands. $\mathbf{T}$ can also be obtained from the sensor spectral radiance responses [67], but that is not always available.

Fig. 3.3 shows the pan-sharpening products of two image pairs, which will be used in Chapter 4 to demonstrate the performance of our proposed road extraction system. The spatial details and spectral integrity are faithfully preserved in the resultant fused image, which has fine spatial resolution and rich spectral response inherited from panchromatic and multispectral images, respectively.

3.2 Binary Road Mask Generation

The first step for image-based road extraction is to properly identify all possible road pixels within a given multispectral image, which could be either the original or pan-sharpened using NNDiffuse. This idenfication does not need to be very precise but it is desirable to include as many candidate road-like pixels as possible. Road likeness is defined as being spectrally or geometrically similar to road samples or templates. Extracted candidate pixels comprise a *binary road mask* (BRM) with all road-like pixels labeled.

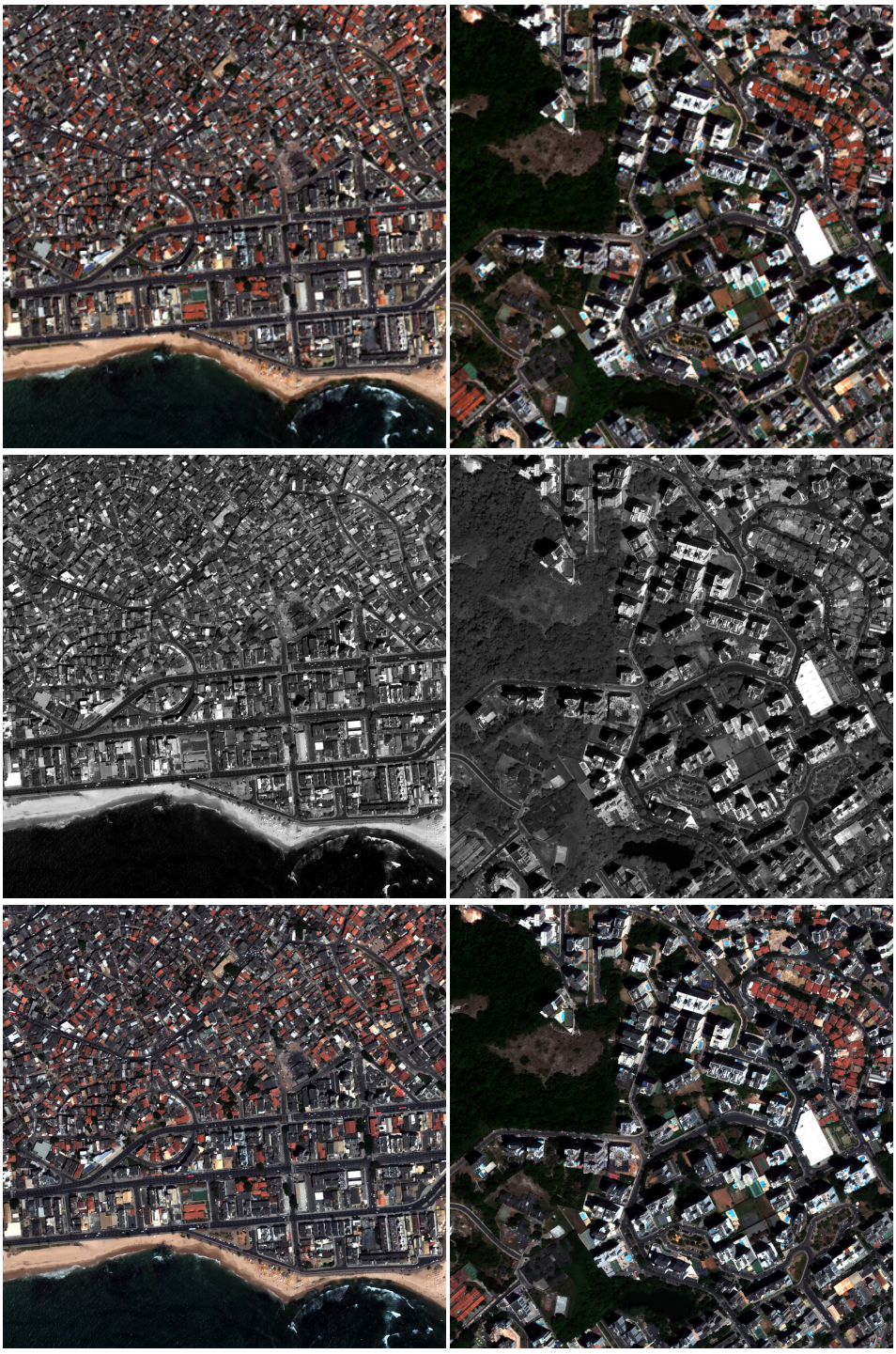


Figure 3.3: *Pan-sharpening of two image pairs. From top to bottom: original multispectral images, panchromatic images, pan-sharpened images. Images details can be found as the 3rd scene in Table 4.1.*

3.2.1 Previous Work

A multitude of methods have been developed to generate a usable BRM for road extraction. A few examples are cited here. [23] used a supervised histogram learning technique to classify on-road/off-road pixels. [21] applied a spatial length-width contextual measure to generate candidate road pixels by setting thresholds for ED, length, and width, which is actually a combination of spectral and spatial features. [15] applied flood filling by jointly using SAM and ED to determine connected asphalt pixels, but their approach cannot extract isolated road regions where no seed point is present. These approaches are not as straightforward as our proposed scheme, which will be described below and has comparable outcomes as the above methods.

3.2.2 Binarized NDVI

One fully automated way to derive a BRM for geographic regions with a vegetated landscape is to compute a binarized NDVI (normalized difference vegetation index) [68] image so as to to mask out vegetated areas. Here NDVI is defined as the normalized intensity difference between near infrared (NIR) band and red band:

$$\mathrm{NDVI}\,(x) = \frac{\mathrm{NIR}\,(x) - \mathrm{red}\,(x)}{\mathrm{NIR}\,(x) + \mathrm{red}\,(x)}, \tag{3.4}$$

where x is one of the image pixels. In addition to its original applications of identifying vegetation and assessing its health, NDVI has proved its effectiveness in masking out vegetated area in road network extraction [21, 69, 70, 71]. A BRM is generated by thresholding NDVI image at τ_{NDVI} according to the following rule:

$$\mathrm{BRM}\,(x) \equiv \mathrm{NDVI}^b\,(x) = \begin{cases} 1 & \mathrm{NDVI}\,(x) < \tau_{\mathrm{NDVI}} \\ -1 & \mathrm{NDVI}\,(x) \geq \tau_{\mathrm{NDVI}} \end{cases}, \tag{3.5}$$

where b represents binarization operator. Road pixels included in the binary mask have the value of 1; rest of the pixels are intentionally mapped to -1 for preparation of the following steps of curvilinear structure detection (Section 3.2) and junction matching (Section 4.1.1).

Obviously, other image objects such as buildings, parking lots, soils, bare rocks, and water bodies are also included in the BRM, which may add to confusion for road separation. To be more specific, despite its applicability to many scenes, binary NDVI is not ideal with images covering the following two scenarios:

1. no vegetation exists along two sides of the road,
2. a dense residential area where the entire road surfaces are barely visible due to tree canopy occlusion and a row of adjacent house complexes immediately along the road are next to each other.

In the first case the roads are completely inseparable from the background, while in the second case adjacent rooftops may be included in the BRM and, worse still, they are likely to look more like road surfaces than the real roads (see Fig. 4.20 for an example). Due to the limitations of binary NDVI, a more generic binary mask generation method will be presented next to mitigate the problem.

3.2.3 Spectral Grouping

To characterize the spectral signatures of the roads in a remotely sensed image, we need to come up with a measure to quantify the spectral variation of road surfaces. This spectral grouping approach is applied instead when the binary NDVI approach is not applicable. A few road pixel samples representing different types and characteristics of road pavements are manually collected from the multispectral image, whether the original or pan-sharpened version. The required number of collected pixels depends on the possible spectral variability of road surfaces, whose materials in an urban setting are primarily asphalt and concrete. More sample pixels are needed if a scene contains very diverse road surfaces.

With weathering and use to varying degrees, roads surfaces paved with the same material may look different over time. In practice, the spectral change would be either brightened, darkened, or even smeared. In some cases, the pavement color (or spectrum) may change as well. To account for these appearance changes, when comparing similar road pixels, variations of both spectral magnitude (brightness) and spectral direction (color) should be taken into account. The spectral similarity measure ATD [72, 73] that was introduced in Chapter 2 is particularly suited for spectral comparison of urban objects with some spectral variations and is defined as:

$$\mathrm{ATD}\left(\mathbf{x},\mathbf{y}\right)=\left(\mathbf{x}-\mathbf{y}\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathbf{x}-\mathbf{y}\right), \tag{3.6}$$

where $\mathbf{x}$ is the target pixel spectrum to be compared and $\mathbf{y}$ is the reference pixel spectrum. This measure is similar to Mahalanobis distance (MD) but its covariance matrix Σ is tunable and is defined as:

$$\Sigma=\mathrm{I}+\left(w^2-1\right)\mathbf{y}\mathbf{y}^{\mathrm{T}}/\left\|\mathbf{y}\right\|^2, \tag{3.7}$$

where I is an identity matrix and w is a shape parameter controlling the measure sensitivity to the change of spectral magnitude or spectral direction. Each one of the pixel samples, $\mathbf{y}_i$, $i=1,2,\cdots,N$, where N is the number of collected sample pixels, serves as a unique reference spectrum. Note here DOS is not required because the spectral variation among road pavements are not due to varied illumination conditions, but rather different materials. The spectral similarity measurements of all image pixel spectra are compared against each individual reference spectrum and are then thresholded to yield separate binary masks. The process is carried out for each $\mathbf{y}_i$ and combined together:

$$\mathrm{BRM}\left(x\right)=\mathrm{ATD}^b\left(\mathbf{x},\mathbf{y}_1\right)\mid\mathrm{ATD}^b\left(\mathbf{x},\mathbf{y}_2\right)\mid\cdots\mid\mathrm{ATD}^b\left(\mathbf{x},\mathbf{y}_N\right), \tag{3.8}$$

where b represents binarization operator, and $\mid$ represents OR operator. The final

BRM is generated with a selected ATD shape parameter w and threshold value τ_{ATD}. The resultant BRM contains only those pixels spectrally similar to either one of these N collected sample pixels. The masked-out (zero-value) pixels in the BRM are then remapped to -1 for the convenience of subsequent processing.

One of the advantages of using ATD as the spectral similarity measure is that in practice only τ_{ATD} needs to be pre-specified, since the selection of w is not critical and can be determined beforehand from a constrained range (empirical value ranges from 3 to 7). However, two threshold values τ_{SAM} and τ_{ED} need to be determined in [15] and the assignments are not intuitive because they are different physical quantities, i.e., angle vs. distance. Additionally, ATD can be freely tuned by adjusting w according to varying scene content, while SAM+ED can only approximate this behavior by changing threshold values. Moreover, if compared with single spectral similarity measures, capturing spectral changes in magnitude and direction gives a more robust result than focusing only on either aspect of SAM or ED. Another benefit is that if some road segments are not included in the BRM due to missing sampling or scene expansion, simply adding one sample pixel on that lost road segment would suffice to address the problem. Finally, ATD provides a human-in-the-loop action but only requires minimum sampling effort. As long as a multispectral image is available, spectral grouping using ATD is straightforward to implement and is sufficiently accurate and robust for BRM generation. It is possible to use RGB images for this method, but lack of NIR band may weaken the spectral separability between vegetation and roads.

Examples of applying ATD to two real images (Fig. 3.3) are shown in Figs. 3.4 and 3.5, respectively. It has been found that $(w, \tau_{\text{ATD}}) = (5, 50)$ is a good combination to yield consistent and acceptable BRMs. Note that the images are captured by WorldView-2 sensor and has 11-bit dynamic range (0 - 2047). Comparisons of spectral gouping using ATD vs. combined SAM and ED are shown in Figs. 3.4e, 3.5e, and 3.6, respectively. ATD yields comparable, if not better, results. It is preferable of choosing ATD in spectral grouping to generate a BRM. Further, one of its parameter w can be readily determined since BRM does not

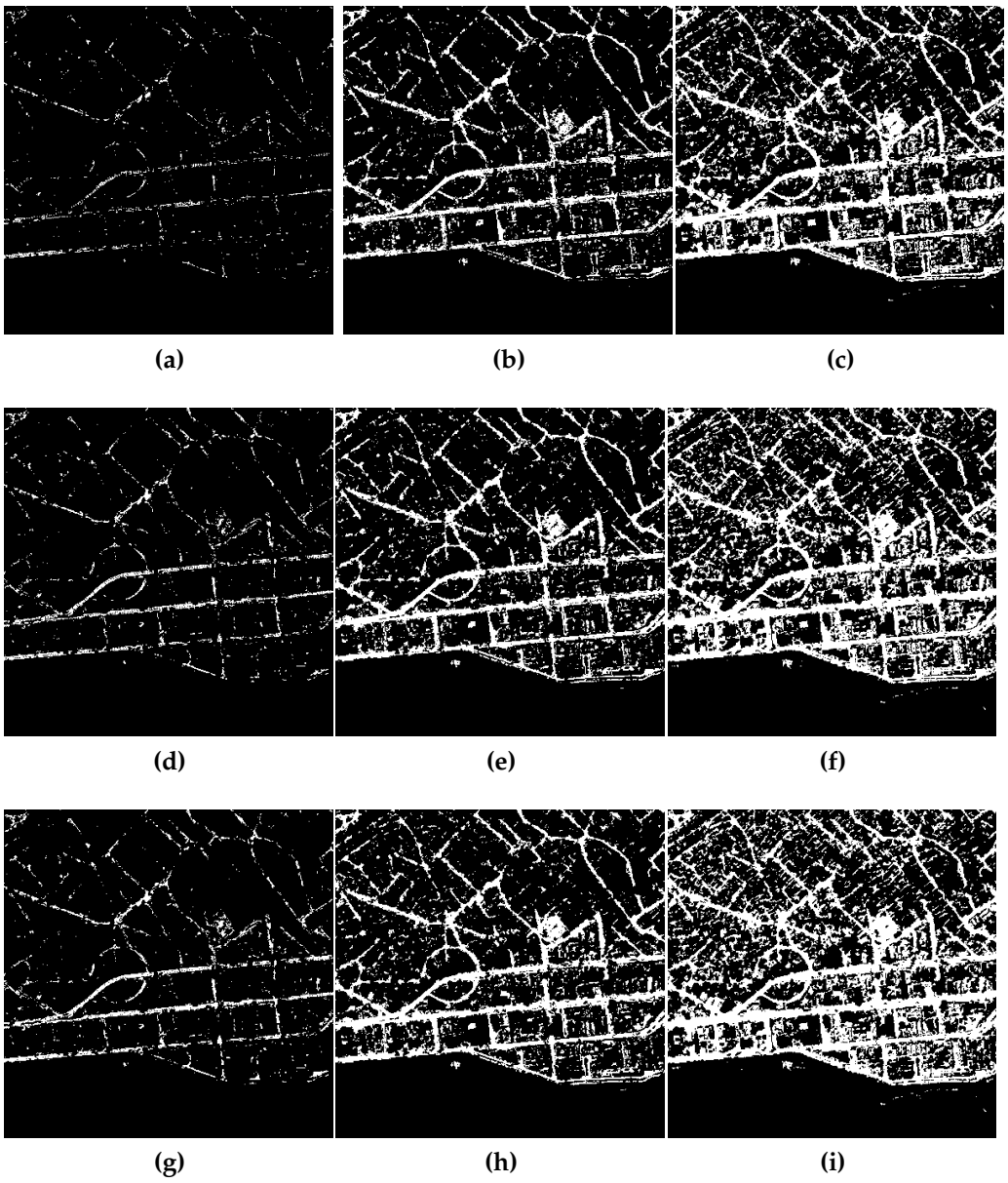


Figure 3.4: *BRM of the 1st image tile with varying ATD parameter w and thresholding value* τ_{ATD}. *(a)* $w = 3$, $\tau_{ATD} = 30$. *(b)* $w = 3$, $\tau_{ATD} = 50$. *(c)* $w = 3$, $\tau_{ATD} = 70$. *(d)* $w = 5$, $\tau_{ATD} = 30$. *(e)* $w = 5$, $\tau_{ATD} = 50$. *(f)* $w = 5$, $\tau_{ATD} = 70$. *(g)* $w = 7$, $\tau_{ATD} = 30$. *(h)* $w = 7$, $\tau_{ATD} = 50$. *(i)* $w = 7$, $\tau_{ATD} = 70$.

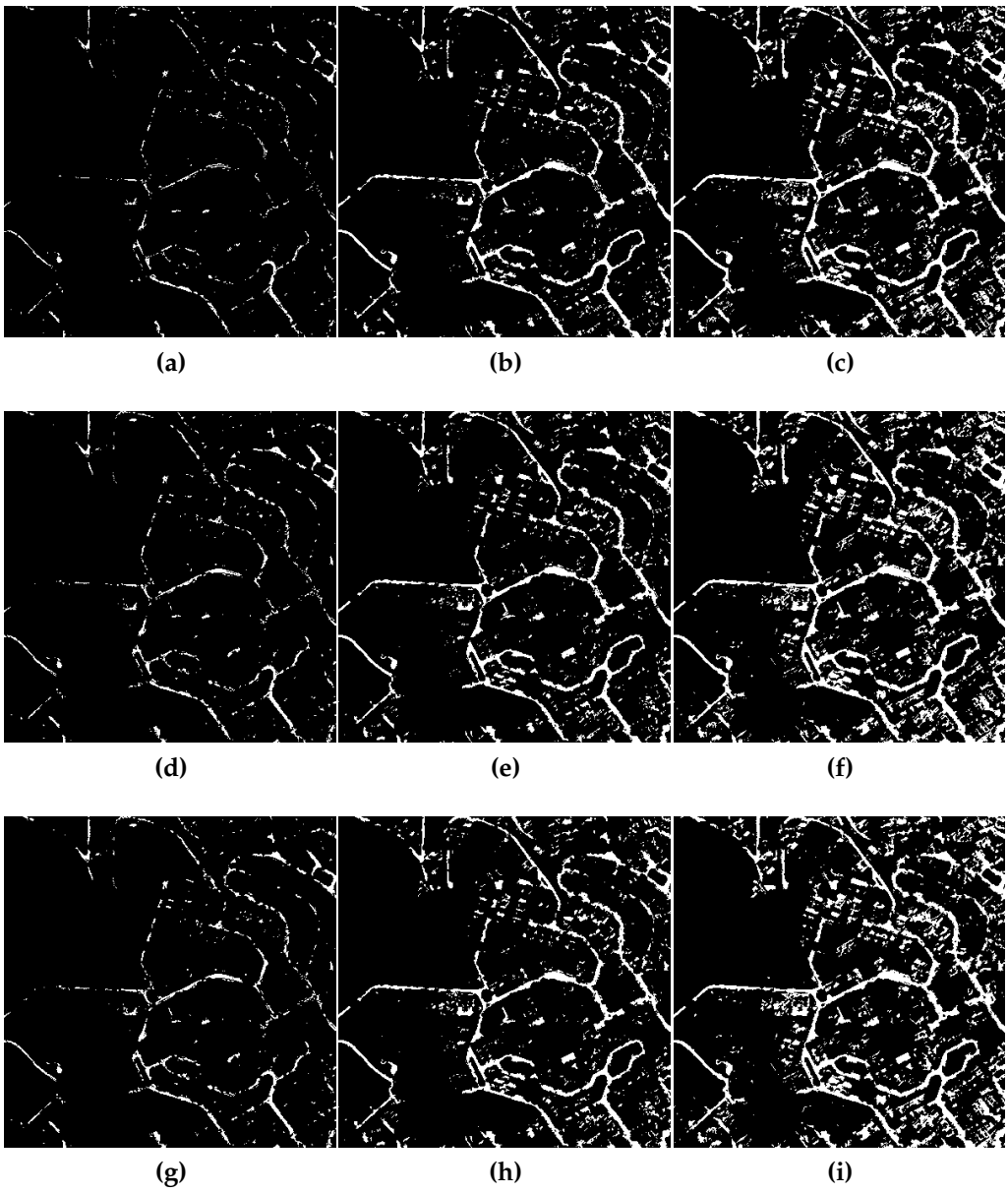


Figure 3.5: *BRM of the 2nd image tile with varying ATD parameter w and thresholding value* τ_{ATD}. *(a)* $w = 3$, $\tau_{ATD} = 30$. *(b)* $w = 3$, $\tau_{ATD} = 50$. *(c)* $w = 3$, $\tau_{ATD} = 70$. *(d)* $w = 5$, $\tau_{ATD} = 30$. *(e)* $w = 5$, $\tau_{ATD} = 50$. *(f)* $w = 5$, $\tau_{ATD} = 70$. *(g)* $w = 7$, $\tau_{ATD} = 30$. *(h)* $w = 7$, $\tau_{ATD} = 50$. *(i)* $w = 7$, $\tau_{ATD} = 70$.

(a) (b)

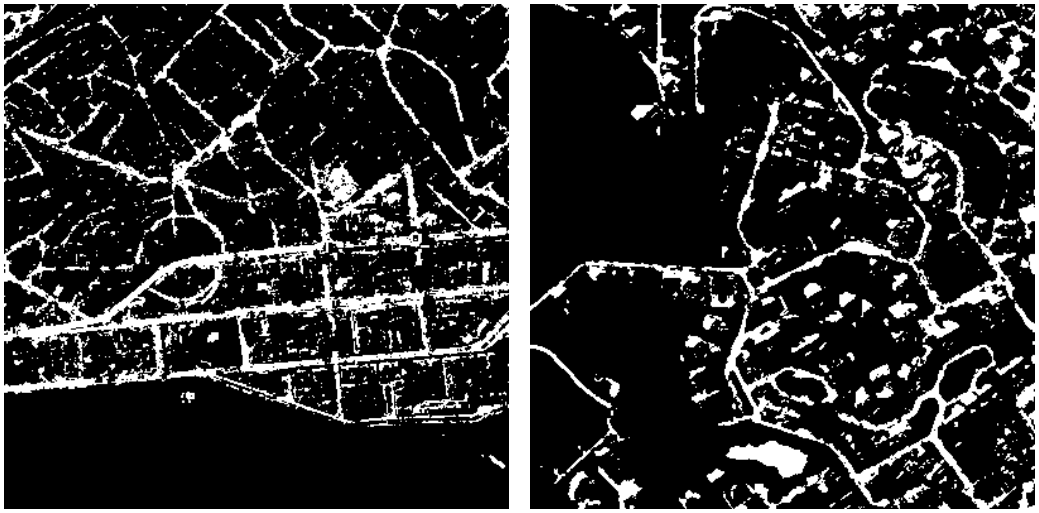
Figure 3.6: *BRMs generated by combing binarized SAM and ED images.* $\tau_{SAM} = 0.1$ *(radian) and* $\tau_{ED} = 100$*. (a) 1st image tile. (b) 2nd image tile.*

change significantly as w changes as shown in Figs. 3.4 and 3.5. Empirically, the optimal range of w is from 3 to 7 and $w = 5$ is chosen as a fixed shape parameter for all experiments in this dissertation. In this way, basically only τ_{ATD} should be changing according to different scenes.

3.2.4 Shadow Pixel Aggregation

It is not uncommon in an urban environment that quite a few road surfaces are entirely or partially cast by shadows. These road pixels in the shadows, however, should be included in the BRM by spectral grouping; otherwise the presence of roads in these areas may become ambiguous. A simple way to exploit the unique capability of WorldView-2 spectrometer (and potentially WorldView-3) is to compute the *shadow pixel index* (SPI) using the coastal blue band and the NIR band:

$$\text{SPI}\,(x) = \frac{\text{coastal blue}\,(x) - \text{NIR}\,(x)}{\text{coastal blue}(x) + \text{NIR}\,(x)}. \tag{3.9}$$



(a) (b)

Figure 3.7: *Shadow pixel aggregated BRMs. (a) and (b) are derived from Figs. 3.4e and 3.5e, respectively, with shadow pixels aggregated.*

This index takes advantage of the fact that stronger atmospheric scattering occurs at shorter wavelength and thus shadow pixels will become prominent in an SPI image. Binary SPI is generated by simple thresholding and is integrated into the regular BRM. Note that water is also evident in a binarized SPI image but it can be easily removed by setting a criteria for the region area: if any of the connected binary regions are larger than a predefined area threshold, it will be identified as water pixels, rather than road candidate pixels. The resultant binary road mask or BRM is aggregated as

$$\mathrm{BRM}\left(x\right) = \mathrm{ATD}^{b}\left(x\right) | \mathrm{SPI}^{b}\left(x\right), \tag{3.10}$$

where b represents binarization operator and $|$ represents OR operator. Certainly, images from other sensors could be used, but carefully choosing shadow pixels as reference road pixels is recommended.

Examples of shadow pixel aggregated BRMs are shown in Fig. 3.7. Note that many broken roads have been bridged by identified shadowed road pixels. To

improve the usability of the BRM, morphological opening is also applied to fill in small holes and breaks. An accurate generation of a BRM is crucial to the performance of the whole system, given that it is the first step of the algorithm pipeline and the following steps (curvilinear feature extraction and junction matching) rely upon its quality. Manual inspection of a BRM is advised to make sure its quality meets the requirements in the following guide. As a general guidance, a good BRM should embody the following characteristics:

1. a clean outline of road topology,

2. faithful representation of junction shape and road width,

3. a reasonably wide buffer zone between roads and other roadside objects along them,

4. no large breaks or holes on any single road surface.

3.3 Curvilinear Structure Detection

Among the most fundamental features of a remotely sensed image are linear structures, since line segments are ubiquitous in many kinds of man-made image objects. An accurate identification of line segments is critical to subsequent recognition and understanding of higher-level objects. Paved roadways captured in orthoimagery can be geometrically modeled as long-thin lines or strips with variable width and orientation. Such feature characterization of roadways makes a curvilinear structure detector well suited for the extraction of road pixels within a typical nadir image. The multi-scale strategy of detection adopted in [74] is to use a set of curvilinear detectors with different widths and orientations to form a series of filter banks. The BRM is then used to convolve with the *filter stack* to obtain a set of curvilinear pixels. One of the detectors takes the form of a long rectangular shape and is defined as:

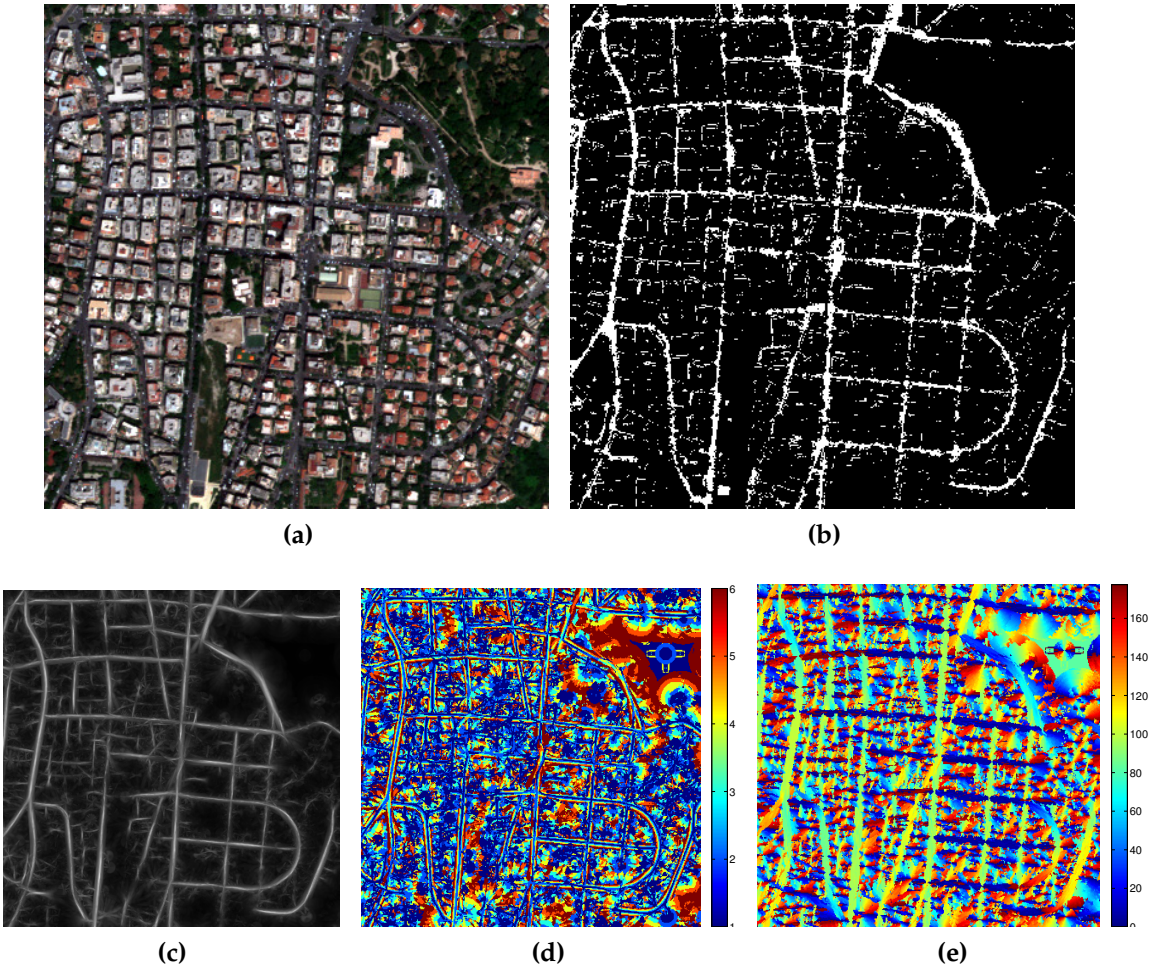


Figure 3.8: *Curvilinear structure detection. (a) Multispectral image. (b) Generated BRM from (a). (c) Maximum projected curvilinear response image. (d) Stack index map of (c). (e) Orientation (in degree) map of (c).*

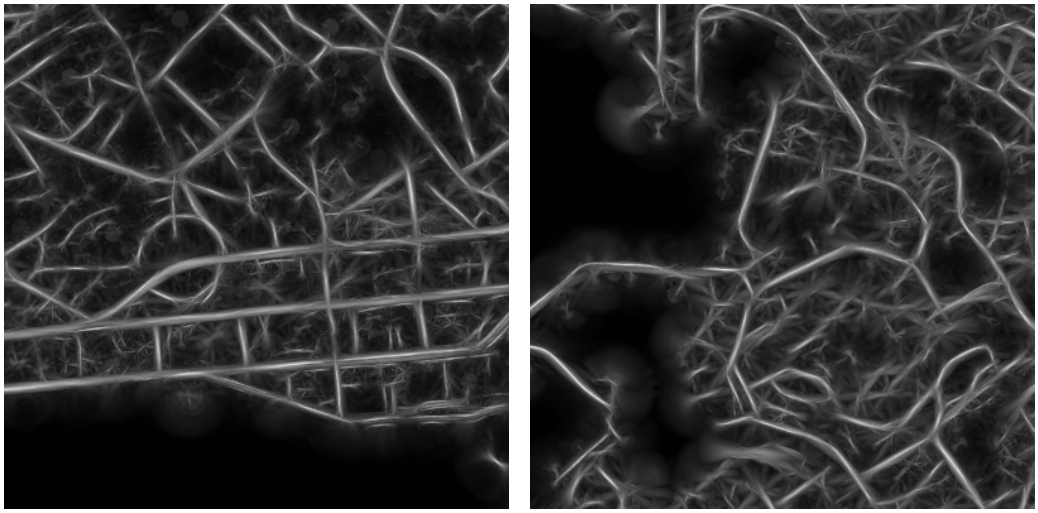
$$f(x;\omega,\theta) = \frac{1}{2\omega r}\begin{cases} 1 & |x| < \frac{\omega}{2} \\ -\frac{\omega}{2l} & \frac{\omega}{2} \leq |x| < \frac{\omega}{2} + r \\ 0 & |x| \geq \frac{\omega}{2} + r \end{cases}, \tag{3.11}$$

where x is one of the pixels in the filter bank, ω is the length of the short rectangle edge and is representative of the road width, θ is an implicit parameter for the orientation angle, r is the boundary width and is fixed, and l is the length of the long edge of the rectangle and is proportional to the value of ω; the region between $\frac{\omega}{2}$ and $\frac{\omega}{2} + r$ is designed to capture the boundary response so that the template gives maximum response when correlated with two parallel edges with road pixels filled in between separated by ω. The generated curvilinear output is given by convolution:

$$c(x;\omega,\theta) = \mathrm{BRM}(x) * f(x;\omega,\theta), \tag{3.12}$$

where $c(x;\omega,\theta)$ is one of the curvilinear response images given a specific combination of width and orientation. The collection of all response images corresponding to varied filter banks can be imagined as a response image stack with the number of stack layer equal to the number of all filter banks. Generally, the largest per-pixel response along the stack layer is computed to obtain a maximum projected *curvilinear response image* and an associated *stack index map* of equal size to indicate which stack layer, or equivalently filter bank, produces the maximum response at that specific pixel location. Pixels with higher curvilinear response values are more likely to belong to the road centerlines. The stack index map is used to indicate the index of the filter bank generating the maximum response along the stack layer direction in any one of the pixels. The filter bank is believed to have a good chance of reconstructing the geometric shape of the road segments within a neighborhood of the pixel of interest.

An example of curvilinear structure detection is given in Fig. 3.8. The BRM is generated from a multispectral image and is then used to create the curvilinear



(a) (b)

Figure 3.9: *Maximum projected curvilinear response image. (a) is derived from the BRM in Fig. 3.7a and (b) is derived from the BRM in Fig. 3.7b.*

response image as shown in Fig. 3.8c. Though irrelevant linear structures are visible too, road structures are cleanly traced as the local response peaks (brightest pixels), which effectively approximate the road centerlines. The associated stack index map and orientation map are byproducts of curvilinear structure detection and will be used to help extract complete road information.

Another two examples are shown in Fig. 3.9 to further demonstrate the detection performances. These curvilinear response images are created based on the BRMs in Fig. 3.7. Though the image background is fairly cluttered, those brightest pixels in localized regions again provide good estimates of road centerlines that will be conducive to road pixel extraction. Curvilinear structure detection results will be used in the ensuing steps of junction matching, intermediate point matching, and road pixel mask generation in Chapter 4.

Many linear feature extraction techniques were applied in the past:, e.g., Hough transform, Radon transform, gradient signature, detections of linear edge [24] or road boundaries [30]. However, the curvilinear structure detector differentiates

itself from other linear feature detectors. The extracted image curvilinear features are very informative that they not only delineate road centerlines, but also provide continuous estimation for width and orientation along the road.

3.4 Vector Road Map Extraction

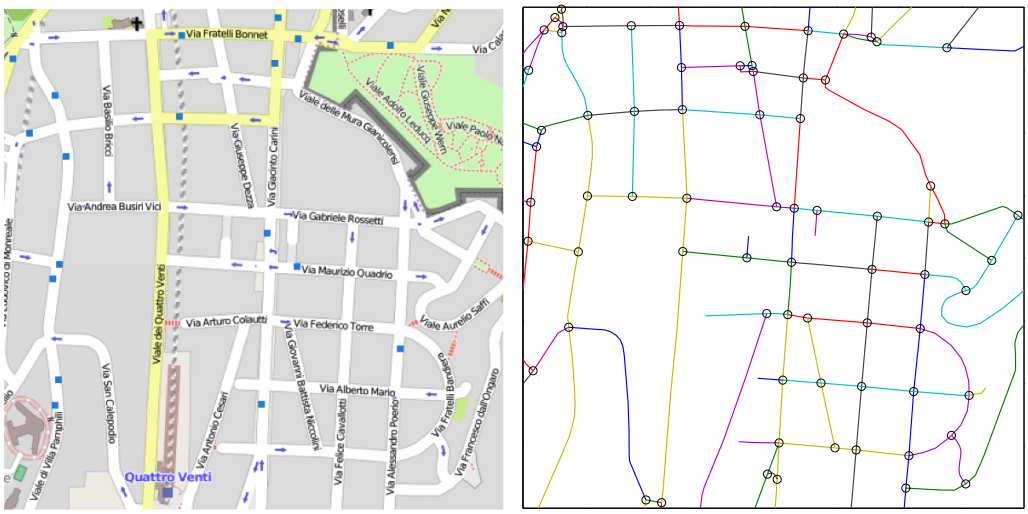
The growth of online map services, especially OpenStreetMap (OSM) [75], leads to easy accessibility of map data and booming application development. OSM (`www.openstreetmap.org`) is a crowdsourcing project created in 2004 with the purpose of building a freely available and editable geographic road network database of the world and meanwhile it is a widely used online free road map provider. Unlike other geospatial data, there are virtually no restrictions on the use and application of OSM data. You can use it for any purpose, including commercial activities, without having to pay license fees. [76] The quality of OSM data has been quantitatively evaluated and is proven to be fairly good, though vector adjustment might still be needed. [77, 78] Thanks to the collaborative and continuous efforts from 2-million active global contributors [79], OSM coverage keeps growing and map quality always improves over time.

The full OSM database and full history dump, in the format of *a planet file*, are free for public access. However, the Esri® *shapefile* format, rather than the original data format, is preferred because the shapefile is a common standard file type for representing geospatial vector data and many geospatial processing tools are able to support, e.g., ArcGIS and QGIS. The OSM data extracts are converted to shapefiles and are categorized by different geographical regions, which are available online for download in Geofabrik [80]. Normally, shapefiles with the following classes are provided: *buildings*, *landuse*, *natural*, *places*, *point*, *railways*, *roads* and *waterways*, among which only roads within the cropped area of the corresponding geo-referenced image are used. A road segment is represented in the format of vectors, or equivalently a combined polyline, which is an efficient way of representing routes on a map. A polyline can have multiple attribute fields, e.g., *lat-*

itude, *longitude*, *street name*, *oneway*, *speed limit*, etc, describing the geographic and physical information of that road segment. The road vectors within the selected region of interest are imported and geographic coordinates are transformed to projective coordinates, as well as image intrinsic coordinates, in accordance with the projection parameters from the metadata of a geo-referenced image. Aside from accessing a road network in original vector format, we rasterize it in accordance with the regular image grid so as to facilitate image-based analysis. In this research, the road type attributes of *motorway*, *primary*, *secondary*, *tertiary*, *residential*, *motorway_link*, *primary_link*, *secondary_link* are selected to screen target road vectors that will be used for map conflation and image road extraction. The generation of geometric templates of road segments using these road vectors and rasterized road vectors will be discussed in Chapter 4.

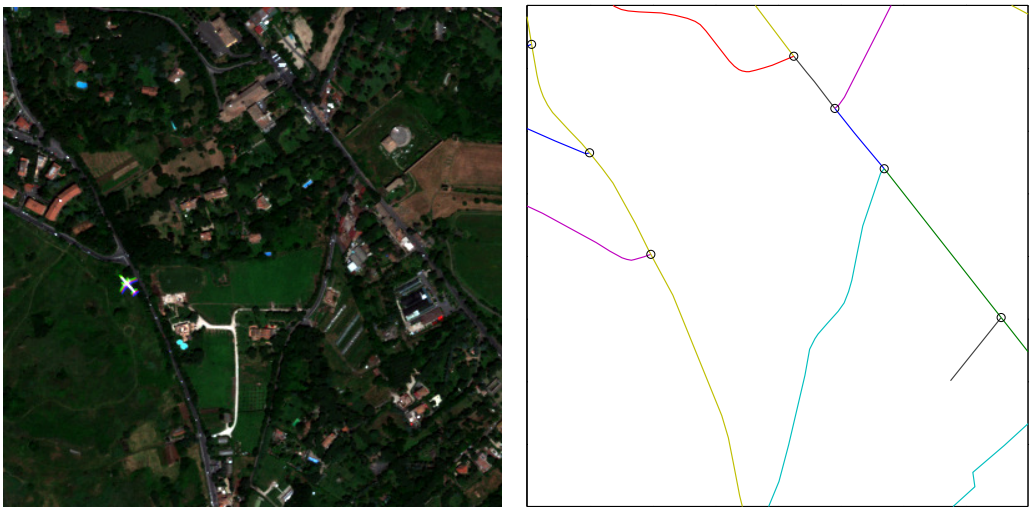
Additionally, in order to account for the inconsistency of data representation, the original road vectors are standardized so that each *segment* has at least a junction point as one of its two endpoints, where *junction point* (also known as intersection point or branch point) is defined as a vector point with more than two connected vertices. In other words, a junction point connects three or more segments together; similarly, an *intermediate point* is a road vector point that is not a junction point. Standardization is necessary because it transforms the map data with different representations into a single framework. An edge linking two connected vertices is called a *fragment*. Suppose there are n vertices within a segment, then there are $n-1$ fragments in this segment; if there is no intermediate vertex within a segment, its fragment is equivalent to the segment.

Derived from the OSM data shown in Fig. 3.10a, the road vectors after standardization are shown in Fig. 3.10b and road segments are drawn with varied colors. Junction points are also labeled as the result of standardization. Referring to Fig. 3.11b, another standardized vector map is also generated as a comparison and its corresponding geo-referenced image is shown in Fig. 3.11a for reference.



(a) **(b)**

Figure 3.10: *Vector road map generation. (a) Online OSM map screenshot. (b) Extracted OSM road vectors. Lines with different colors correspond to different segment. Circles highlight extracted junction points.*



(a) **(b)**

Figure 3.11: *Vector road map generation. (a) Geo-referenced multispectral image. (b) OSM road vectors. Lines with different colors correspond to different segment. Circles highlight extracted junction points.*

3.5 Summary

In Chapter 3, representative road features from a geo-referenced image and a digital road map are created as the basis for the ensuing road map conflation and image-based road extraction. If necessary, a multispectral image is pan-sharpened by using NNDiffuse to yield a resolution enhanced multispectral image for the purpose of more precise discrimination on road pixels; A binary road mask is generated from binary NDVI or spectral grouping and is used to create a corresponding curvilinear structure response image, and altogether will be applied in feature matching, which will be discussed in Chapter 4. Road map data is converted to compatible vector graph and rasterized for image analysis.

The purpose of road feature extraction is to exploit representative features that can be applied to unambiguously determine the presence and characteristics of roads in the image. Road features commonly used involve two aspects: spectral and spatial. The spatial road feature is often referred to as road geometry or shape. Combination of both aspects provides robust and rich road features. A binary road mask is generated by extracting spectral features from the multispectral image. It also provides critical information regarding the road shape. A curvilinear response image is created by taking advantage of such spatial features to shed light on the unique road geometries. Road vectors are another form of spatial features and will be used to create geometric road templates. The road features generated in this chapter will be passed on to the building blocks within the lower portion of the algorithm pipeline in Fig. 3.1. Since their qualities (in terms of feature accuracy and richness) are critical to the success of road network extraction, those methods to extract these features must be sufficiently robust and informative if they are to serve our purpose well.

Chapter 4

Map Conflation and Image-Based Road Extraction

The road features generated in Chapter 3 will be used in map conflation and road extraction described in this chapter. System workflow is shown in Fig. 4.1 and its bounded portion of the road extraction stage will be discussed. Those enclosed building blocks represent the steps and outputs of automated map conflation and image-based road extraction. Some further background is provided in Section 4.1. The detailed description of our proposed approach is given in Section 4.2 and experimental results on real image data are presented in Section 4.3. Finally, this chapter is summarized in Section 4.4.

4.1 Methodology

The basic design of the road extraction system is to make use of multispectral imagery in conjunction with digital road map data. Since a satellite or aerial image is generally geo-referenced, meaning that it is projected onto a standard cartographic coordinate, e.g., Universal Transverse Mercator (UTM), a road map can be simply geo-registered onto the image plane as long as they share the same projected coordinate system. However, due to a variety of errors, such as sur-

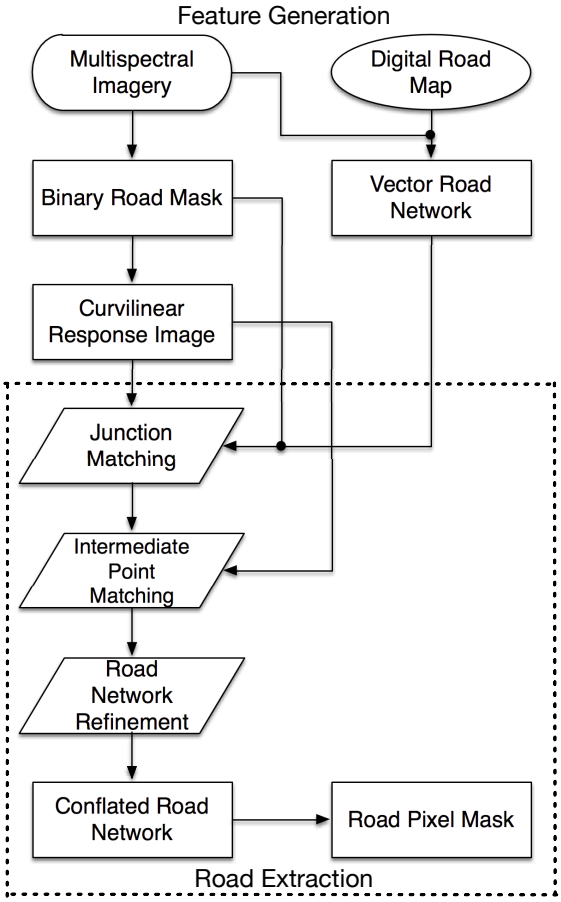


Figure 4.1: *System workflow for map conflation and image road extraction. The road extraction stage is the focus of this chapter. Road features extracted from the previous feature generation stage are passed into the road extraction stage as indicated by the bounding box.*

vey inaccuracies and topographic relief, the road network vectors usually do not align with the road centerlines in the geo-referenced image. A nonsystematic correction is often required to align the road vectors with the roads in the images because the misalignment cannot be compensated by a constant shift. At this point the conflated road vectors can be leveraged to extract road features from the image. Here the road extraction result differs from some prior work where only image road centerlines are extracted. Our approach, however, attempts to exploit the image to extract not only the road centerline, but also road width and orientation, connectivity, and even topology. With this all-around road information as output, the road extraction method is more likely to meet the requirements of real-world applications.

Introducing a road map significantly reduces the detection ambiguities that arise among techniques that do not use prior road knowledge, though in this case registration errors must be fixed first. A typical image registration procedure consists of four basic steps: feature detection, feature matching, mapping function design, and image transformation and resampling [81]. Image-derived features usually consist of a set of extracted control points (tie points). If the control points are geographically referenced, this procedure can further be called geo-registration. In our research this has been extended to the conflation of heterogeneous geospatial data sources: a geo-referenced image and a geo-referenced digital road map. Depending on the geodetic accuracy of the data and the ultimate application, either source can be selected as the reference. For example, [21, 23] use the image as the reference, while [24] uses the road map. Note that in the former case the last two steps of image registration are usually unnecessary. Because of the final goal of our research is to extract road pixels, the image is used as our reference data.

Referring to Fig. 4.1, a brief description of the road extraction stage is as follows:

1. The map-derived vector road network is utilized to construct junction templates for local matching with image features - a BRM and a curvilinear

response image - to determine the locations and geometries of road junctions in the image. Junction points are then corrected in accordance to the calculated image locations and intermediate points are partially corrected based on junction alignment.

2. Junction-corrected intermediate points are further conflated according to local peaks of curvilinear response image via transversal search.

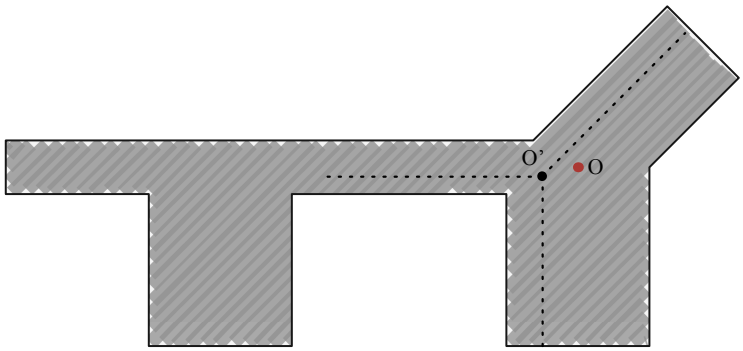
3. Conflated road network vectors are refined and then used to extract a road pixel mask.

4.1.1 Junction Matching

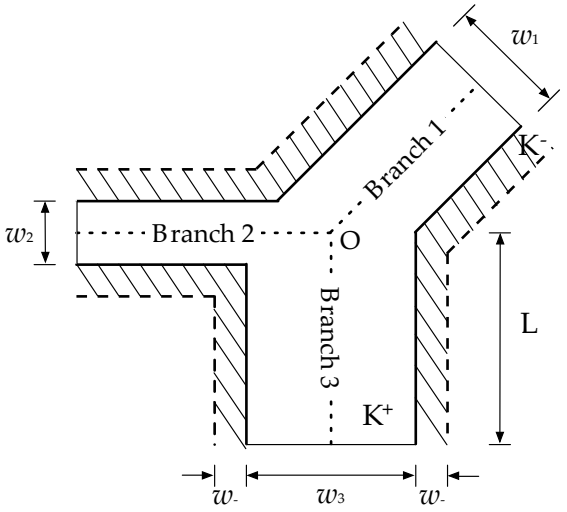
Accurate recognition of road junctions in the image ensures intact road connectivity and faithful recovery of the road network. Junctions are often extracted first since their shapes (3-way, 4-way, or multi-ways) are relatively unique in a localized region and thus can be robustly identified. Local template matching is used in [23] to find control point pairs of road network intersections and image road intersections. A template is generated from vector road data whereas the junction branch width is directly derived from metadata. However, the metadata may be unavailable, e.g., OpenStreetMap, or unreliable. We propose an improved template matching approach to rectify misaligned junctions and identify road branch width by comparing a series of junction templates against the BRM. Junction matching describes the steps taken to find the correspondence between a map junction and its image junction counterpart.

4.1.1.1 Junction Template Generation

A junction template is created based on junction geometries derived from the road map and formed by varying combinations of branch widths properly sized according to the image spatial resolution. Among all shape hypotheses, we want to find the the most likely junction template that matches best with the image



(a)



(b)

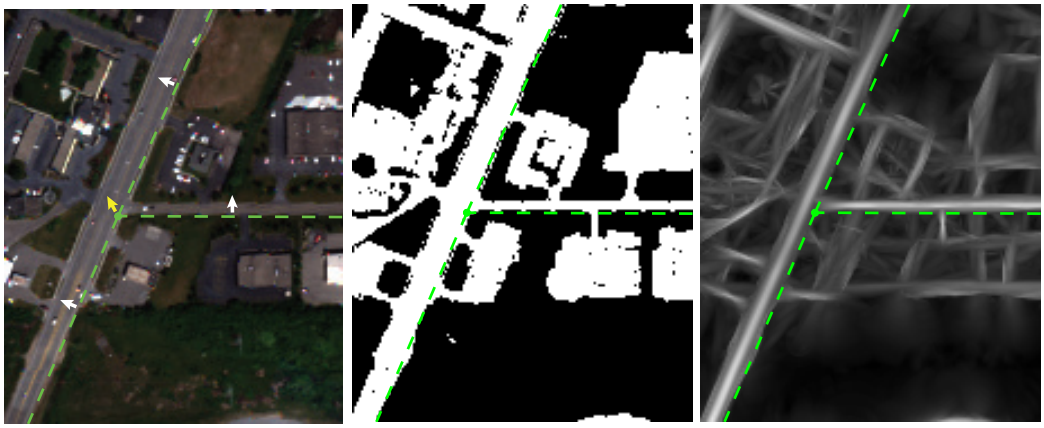
Figure 4.2: *Creation of a three-way junction template. (a) A BRM overlaid with geo-registered initial junction vectors to show the misalignment. The shaded area represents the possible road pavement and is mapped to +1, while elsewhere to -1. The image junction is indicated by O and the map junction is indicated by O'. (b) A three-way junction template with varied branch width corresponding to the initial junction vector. O indicates the junction point. Dotted lines demonstrate the skeleton of the junction. The white region bounded by solid lines indicates positive area K^+ used for junction branch matching. The disjoint textured region shows negative areas K^- used for matching with boundary pixels. The branch widths of three branches are w_1, w_2, and w_3, respectively. The boundary width is equal to w_- on one side and is the same for all branches. L is equal to the junction range.*

junction. This matching scheme is similar to the multi-scale strategy employed in curvilinear structure detection (see Section 3.3). Based on the junction vectors shown in Fig. 4.2a, a 3-way junction template is created as shown in Fig. 4.2b. Area K^+ matches with potential road pixels and disjoint areas K^- correspond to pixels do not appear to be road-like. K^+ is created by expanding each individual junction branch from the road skeleton vectors to the variable full width $\mathbf{w}_+$, where $\mathbf{w}_+ = [w_1, w_2, \cdots, w_N]^{\mathrm{T}}$ and N is the number of junction branches. K^- is constructed by expanding an additional fixed amount of w_- starting from the boundary of K^+. The general junction filter is defined as:

$$g(x; \mathbf{w}_+, w_-) = \frac{1}{2} \begin{cases} \frac{1}{\text{Area}(K^+)} & x \in K^+(\mathbf{w}_+) \\ -\frac{1}{\text{Area}(K^-)} & x \in K^-(w_-) \\ 0 & otherwise \end{cases}, \tag{4.1}$$

where x is one of the filter pixels. $\text{Area}(K^+)$ is the area of K^+ and is approximately equal to $\Sigma_i^N w_i \times L$, where L is the junction range. $\text{Area}(K^-)$ is the total area of disjoint K^- and is approximately $2\Sigma_i^N w_- \times L$. The approximation is made because junction branches have mutual overlaps in the vicinity of the junction point O. In practice, $\text{Area}(\cdot)$ is counted as the total number of pixels. By varying $\mathbf{w}_+$ and fixing w_-, a filter stack encompassing all possible filter banks is generated, which will then be used to match with a locally cropped BRM. Note this filter (Eq. 4.1) is zero-mean and normalized, which means the matching response across different filters can be fairly compared. This is necessary since we need to determine the most likely template among the whole filter stack.

A real example is given in Fig. 4.3a. The original 3-way junction vectors are overlaid on the image by geo-registration and its misalignment with the image road centerlines is evident. The template filter stack derived from the map junction vectors is shown in Fig. 4.4. Each junction template within the stack is generated by rasterizing junction vectors and morphologically dilating road branches with varied widths. Note that in order to reduce the overhead of creating a large number of redundant filter banks, aligned road branches are reasonably assumed



(a) (b) (c)

Figure 4.3: *(a) Cropped junction area of RGB image in Fig. 4.13a. (b) Binary NDVI image of (a) with darkened vegetated areas. (c) Curvilinear structure response image of (a). Brighter pixels indicate stronger response to curvilinear structures. Initial road vectors represented by dashed lines are overlaid on all three images to show the misalignment relative to image road centerlines.*

to share identical width. The template filter stack is inclusive of a range of map junction templates that could match with the counterpart image junction based on knowledge of pixel GSD and expected physical road widths.

4.1.1.2 Local Template Matching

The simple idea of template matching is to find first the single junction template matches best with the image junction and second the image position where the best match occurs. The matching process on the BRM shown in Fig. 4.2a is graphically illustrated in Fig. 4.5. It is easy to see that Fig. 4.5(c) gives the best (exact) match case and the image junction location and shape are readily identified. Cross-correlation is adopted as the matching criterion and is given by

$$c\left(x;\mathbf{w}_{+},w_{-}\right)=f\left(x\right)\star g\left(x;\mathbf{w}_{+},w_{-}\right), \tag{4.2}$$

where $\star$ represents correlation operation. $c\left(x;\mathbf{w}_{+},w_{-}\right)$ is the cross-correlation map for junction matching given one set of combinations of branch width $\mathbf{w}_{+}$

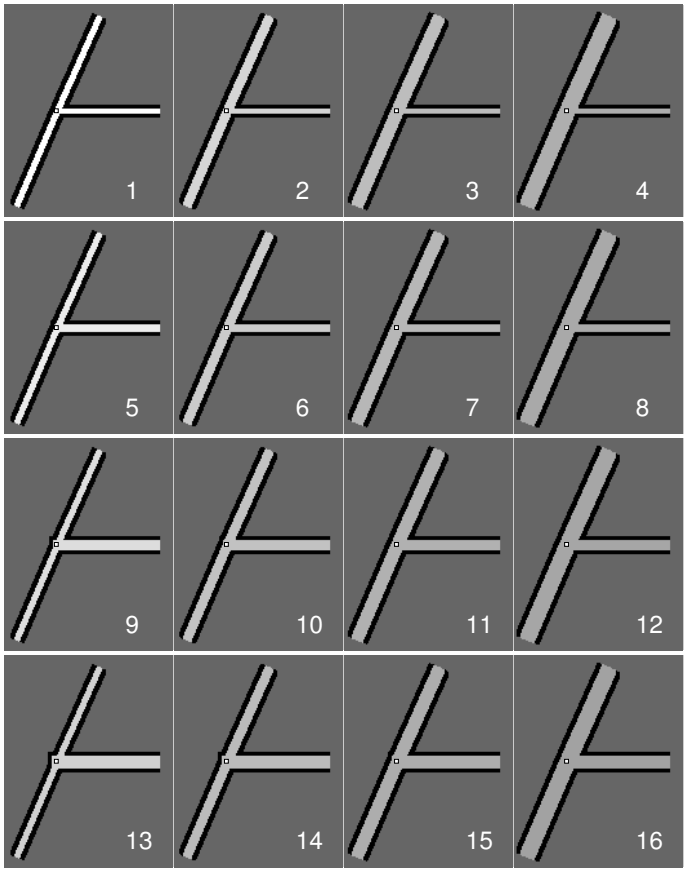


Figure 4.4: *A template filter stack comprising of 16 filter banks. Brighter area corresponds to K^+ and darker area corresponds to K^-. The square symbol indicates the junction point location. Stack index is displayed in each corresponding subfigure. Note that the two vertical branches have the same width since they are assumed aligned. Hence there are equivalently two unique branches, and the total number of filter bank combinations, or stack layers, is 16 ($= 2^4$), given that the number of possible width option of 4, 6, 8, and 10 pixels, which corresponds to physical width ranging from 8 to 20 meters.*

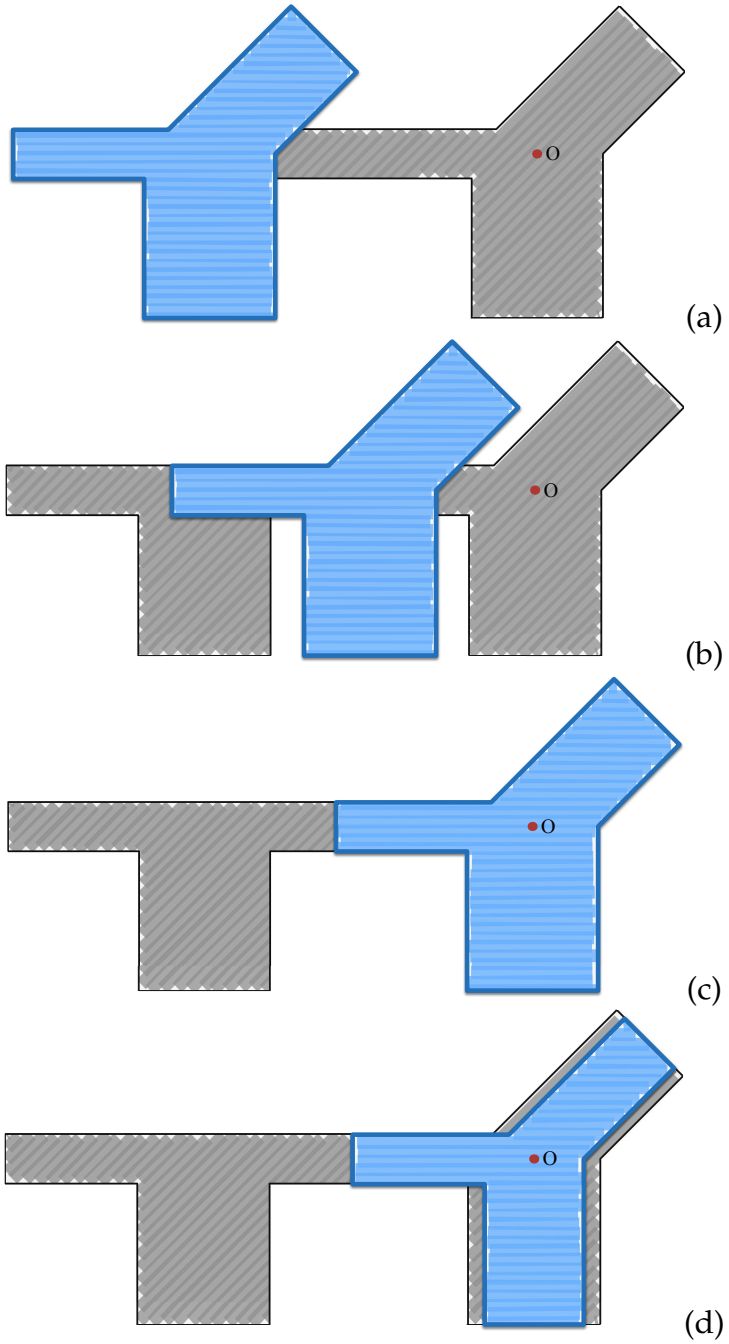
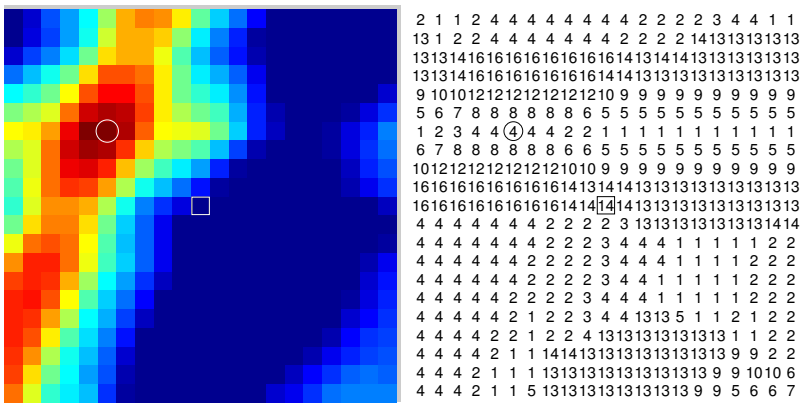


Figure 4.5: *Illustration of local template matching. Shaded area represents the possible road pavement; blue area represents the K^+ area of matching junction template and the surrounding K^- areas have been ignored. From (a) to (c), the template move right and the best match is identified as the graph with perfect match in (c). (d) demonstrates the effect of varied branch width; the template shown is not a good match because it does not fit the image junction as well as the template in (c) does.*



(a) (b)

Figure 4.6: *(a) Maximum projected cross-correlation using BRM. (b) Corresponding stack index map of (a). The square indicates the map center and the circle indicates the found location of best match, in this case the maximum correlation coefficient.*

and $f(x)$ is the locally cropped BRM. Similar to curvilinear structure extraction, a cross-correlation stack is generated using the junction filter stack. This cross-correlation stack is then projected along its layer direction to yield a maximum projected cross-correlation map and an associated *stack index map*. By locating the largest correlation value the image junction location can be determined. Further making a query to the accompanied stack index map retrieves the template that matches best with the image junction and thus defining the widths of the road branches.

Given that the misalignment of vector-to-image geo-registration is generally not very large and is upper bounded, it is unnecessary to search the entire BRM for a possible match. *Local template matching* within an expanded neighborhood of initial image location of the map junction is sufficient and advised for the sake of computational efficiency and algorithm robustness. More beneficially, since one-to-one matching is almost guaranteed within a local region, no feature correspondence matching is necessary, which otherwise was used in [21, 28]. The local search range is normally set to be equal to the largest possible misalignment of map junction and image junction pairs.

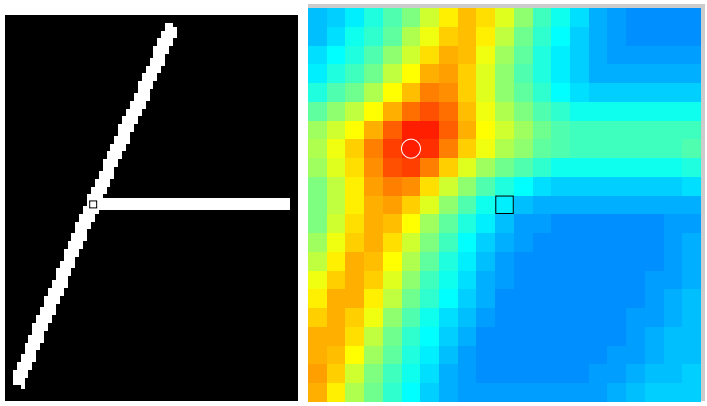
An example of local template stack matching applied on the BRM in Fig. 4.3b to align the map junction is shown in Fig. 4.6. The maximum projected cross-correlation along the stack layer is shown in Fig. 4.6a and the associated stack index map is shown in Fig. 4.6b. The pixel distance between the position of best match (circle) and the search center (square) is assigned as the correction to be imposed on the target junction. The best match position corresponds to the location of the image junction and search center is the initially registered location of the map junction. It can subsequently be inferred from the stack index map that the filter stack index is 4, which means the 4th filter bank (K^+ of that template) of the filter stack shown in Fig. 4.4 bears the most resemblance to the image junction. This can also be verified by referring to Fig. 4.3a. Not only the image junction location is identified (so is the correction amount of the map junction), but its shape is also retrieved in this integrated step.

4.1.1.3 Robustness Enhancement

Since quite a few roadside buildings and parking lots are spectrally similar to the road segments, the non-road pixels may be heavily mixed in a BRM, adding potential confusion to locating image junctions. To address this problem, a complementary matching scheme is applied on a maximum projected curvilinear structure response image, which also gives an estimate of the image junction by matching linear structures of its branches. The spatial filter or template is constructed with only one width option by setting $\mathbf{w}_+$ as a constant value of 3 pixels for all branches and setting w_- equal 0 such that

$$g\left(x; \mathbf{w}_+ = \mathbf{3}, w_- = 0\right) = \begin{cases} \frac{1}{\text{Area}(K^+)} & x \in K^+\left(\mathbf{w}_+\right) \\ 0 & \textit{otherwise} \end{cases} . \tag{4.3}$$

A buffer size of 3 pixels is selected to account for presumably small shape discrepancy between the map junction and the corresponding image junction. The curvilinear response intensities within K^+ are summed in a template matching operation using Eq. 4.1. Note that this filter is no longer zero-mean but is still



(a) (b)

Figure 4.7: *(a) Template correlate with curvilinear response image. (b) Cross-correlation map using curvilinear response image. Square indicates the map center and circle indicates the location of the best match. For the sake of fair comparison, the intensity range is scaled to be consistent with that of Fig. 4.6a.*

normalized. It is worth mentioning that the curvilinear structure detector does not respond reliably in the vicinity of a junction since the junction does not conform to the long-thin rectangular shape hypothesis. To enhance the robustness of the junction matching, both the BRM and the curvilinear response image are utilized in junction matching. The map junction correction is determined jointly by comparing the maximum correlation responses generated by both input images and using the offset position corresponding to the one with larger correlation value.

The template used to match with curvilinear response image (Fig. 4.3c) is shown in in Fig. 4.7a. All three branches have the same width of 3 pixels and there is no K^- area in this template. Similarly, in Fig. 4.7b the maximum position is circled in cross-correlation map generated by correlating the junction template with the curvilinear response image. Since the maximum correlation value by using the BRM is larger than that using the curvilinear response image, the image junction point is then determined by the former.

Moreover, validity checks are imposed for all potential junction pairs to make

sure only those computed correction offsets whose confidence levels are high enough are kept. Two validity rules are set: a strong correlation value and a small relative offset compared with nearby junctions. If either rule is broken, the junction point correction will be invalidated and will be later interpolated based on those that pass the validity test. This check is of great importance to achieve reliable road extraction, because the correction amounts of those junction points that fail the validity check are likely to be erroneous. Should possibly inaccurate corrections be accepted, these errors would be passed onto the next step and deteriorate the detection performance.

4.1.2 Intermediate Point Matching

Junction matching step conflates all map junctions to corresponding image junction locations. In the following step, the techniques used for conflation of remaining intermediate points will be described next.

4.1.2.1 Junction Correction

Aligned junction points essentially serve as control points for map conflation. Intermediate point corrections can be assigned by interpolation based on already corrected junction points. The correction amount assigned to an intermediate point normally moves its corresponding road segment closer to the image road centerline. This procedure is called *junction correction* and is analogues to rubber-sheeting transform for vector-to-imagery conflation in [21, 23] and thin-plate-spline transform for imagery-to-vector conflation in [24], which also rely on control points correspondences. However, neither the above transforms nor the junction correction complete the road map conflation, because residual errors remain.

Junction correction is demonstrated using interchange ramps as shown in Fig. 4.8a. Referring to Fig. 4.8b, the connected local response peaks of curvilinear structures approximately delineate road centerlines. However, the initial road segments are off the centerlines by some varying amount and a translational cor-

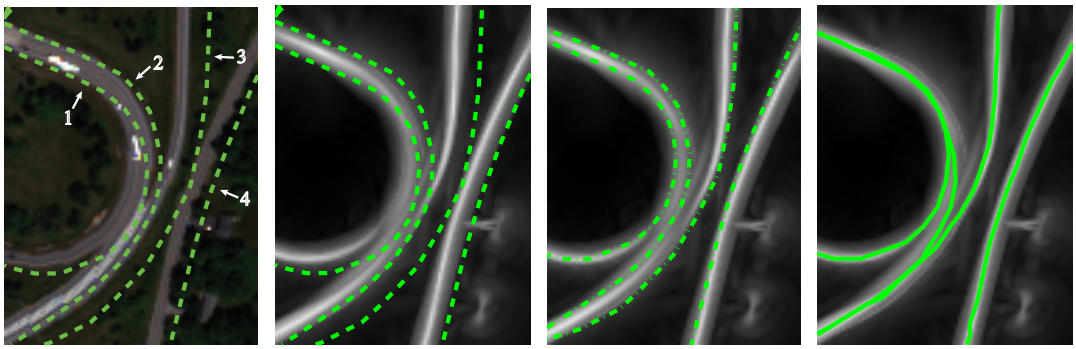


Figure 4.8: *Map conflation. (a) Cropped RGB image from Fig. 4.16a overlaid with initial road segments. (b) Curvilinear structure response image overlaid with initial road segments. (c) Curvilinear structure response image overlaid with junction-corrected road segments. (d) Curvilinear structure response image overlaid with conflated road segments. Brighter pixels in curvilinear response image are more likely to belong to the curvilinear structures. Corresponding binary NDVI image is thresholded at* $\tau_{NDVI} = 0.25$.

rection is not sufficient. Note road segment #3 even overlaps the road surface corresponding to segment #4. Fortunately, junction-corrected road segments based on already conflated junction points (not shown in the figure) shift the initial segments towards their image centerlines as shown in Fig. 4.8c. Junction correction, however, is not uniform. For example, segment #4 is so close to corresponding local curvilinear peaks that only minor adjustment is needed. On the other hand, although segment #3 is much closer to its correct position than before, it still requires further correction.

4.1.2.2 Transversal Search

The junction correction step moves road vectors closer to the image road centerlines but local adjustments of intermediate points are still needed (see Fig. 4.8c). Local template matching is not possible for intermediate point correction because templates generated from these connected vertices lack local geometrical uniqueness, unlike a typical junction. Fortunately, curvilinear structures afford a fairly

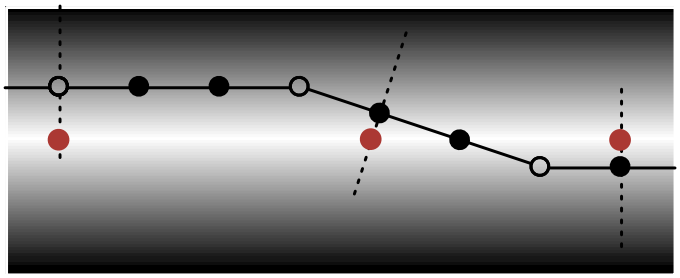


Figure 4.9: *Intermediate point matching by transversal searches in a curvilinear structure response image. The searches are conducted along the dashed lines perpendicular to the directions of road vectors. Brighter area has a stronger probability of the presences of linear structures. Hollow points represent geo-registered original road vector points; solid points represent densified vector points based on original vector points on two ends; red points indicate the matched position of the transversal search of the selected intermediate point.*

good estimate of the road centerlines in the image and can help to rectify map intermediate points.

Referring to Fig. 4.9, by conducting a local *transversal search* along the direction perpendicular to the road vectors in a maximum projected curvilinear response image, the correction amount of a junction-corrected intermediate point is computed as the pixel distance relative to the curvilinear pixel with local response peak along the search line, given that the connected local peaks correspond to the sought image road centerlines. The benefit of this lateral operation is to reduce a 2-D search into a localized search along a 1-D direction. A subsequent query can be made to the stack index map of curvilinear response at the matched local peak to retrieve a width estimation of that road segment portion. Similar to the junction matching, intermediate point conflation and road width retrieval are also naturally integrated: the geometry (orientation) and the width of image road segments are extracted simultaneously.

The prior step of junction correction is always desired because it reduces the effective search range and mitigates the confusion caused by multiple profile peaks along a search line due to the proximity of several road segments. If the transversal search is conducted in Fig. 4.8b, segment #3 is very likely to conflate to

the wrong road. In addition, a long-range search must be performed because this road segment is quite far from the corresponding peak response pixels. Upon completion of junction correction and transversal search, the conflated road segments successfully align with the road centerlines (response peaks) as shown in Fig. 4.8d. Note lack of the median barrier (e.g., vegetation) spectrally dissimilar to the road pavement in between to separate opposite ways of the interchange ramp results in merged response peaks of curvilinear detection. That ramp does not generate two parallel peaks, but a single profile peak that looks just like they are generated from a single wider road. Consequently, road segments are partially merged for #1 and #2, as well as #2 and #3, though the attributes of the individual road segment are preserved.

4.1.2.3 Robustness Enhancement

Validity checks are imposed on every match point to minimize false matches due to a weak curvilinear response or nearby spurious linear structures. Only if 1) the curvilinear response magnitude is larger than a predefined threshold, 2) the curvilinear orientation is compatible with that of the connected road fragments up to a tolerant angle, and 3) the offsets relative to the adjacent points are below a predefined threshold, then the match with that found curvilinear pixel is accepted. Through validity checks many intermediate points are left unassigned, which will be interpolated based on nearby points with valid corrections. And in this way, severe occlusions are naturally handled in our system workflow.

When the curvilinear responses within the search ranges of a number of vertices happen to be very weak due to severe occlusions, too many vertices are left with invalidated corrections. To overcome the difficulty, segment vertices are densified along fragments at a predefined interval to increase the odds of finding valid matches. Densified road vectors are represented by solid points in Fig. 4.9. Provided that some added points, together with the original vertices, produce valid matches, the remaining points with invalid corrections in the same segment can be interpolated with a higher confidence. Moreover, in some extreme but

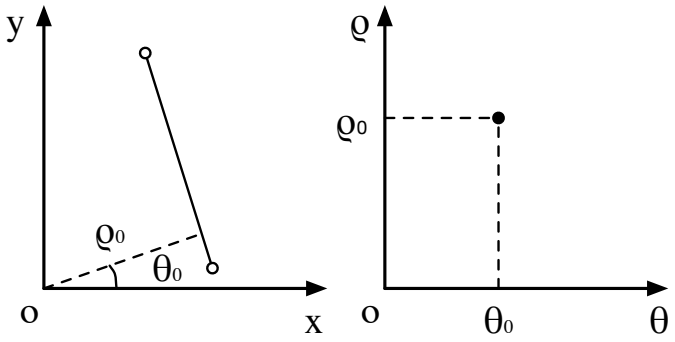


Figure 4.10: *Hough transform. The parameter ρ_0 represents the algebraic distance between the line and the origin, while θ_0 is the angle of the vector from the origin to this closest point. A line represented by (ρ_0, θ_0) is transformed into a point (ρ_0, θ_0) in Hough space.*

not necessarily rare cases, when no valid match can be found for all intermediate points within a segment, the final conflated road segment will be equivalent to the junction-corrected segment and its width will be conservatively assigned with the minimum road width.

4.1.3 Road Network Refinement

In practice, many connected road segments are actually aligned and form a longer straight line. Nevertheless, intermediate point matching based on densified points tends to wiggle an originally straight road. In a refinement step, aligned road segments are stipulated to be collinear. First, each individual road fragment is transformed to Hough space and becomes a feature point as illustrated in Fig. 4.10 [82]. Fully automated mean-shift clustering [83] is adopted to find all possibly aligned fragments either across different segments or within a same segment. An example is given in Fig. 4.11 to show the transform and clustering of aligned road fragments. However, many parallel road segments, which are close to each other but not necessarily connected, are also grouped in clusters. A connectivity check of road fragments is conducted to screen out both connected and aligned road fragments. Finally, least-square linear fitting is employed on grouped frag-

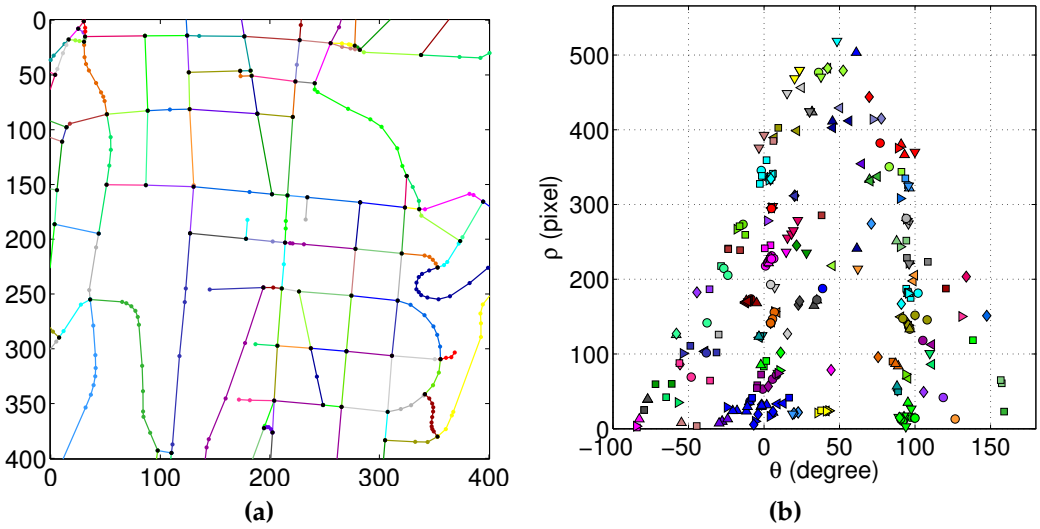


Figure 4.11: *Mean-shift clustering of aligned road fragments. (a) Vector road network. Different colors represent different road segments. Black points represent junction points and the colored points represent intermediate points. (b) Hough space representation of the road fragments in (a). Different symbols correspond to different road segments. Each cluster of symbols with the same color represents the candidates of aligned road fragments.*

	Size (pixel)	Image Type	GSD	Location	Date	Sensor
1	1033×2048	Multispectral	2m	Henrietta, NY	Jun 2010	WorldView-2
2	2048×2048	Multispectral	2m	Rochester, NY	Sept 2010	WorldView-2
3	4040×4200	Pan-sharp	0.5m	Salvador, Brazil	Dec 2009	WorldView-2
4	2798×3300	Multispectral	2m	Rome, Italy	Jun 2010	GeoEye-1

Table 4.1: *Data sheet of test image scenes.*

ments and the image locations of the vector points are updated accordingly. This step concludes the conflation of vector road map to geo-referenced imagery and yields the extracted image road centerlines.

4.1.4 Image Road Extraction

The output of the previous step is the conflated road segments that are equivalent to extracted image road centerlines. Most other road extraction approaches stop there. Our proposed system has an additional step to fully extract knowledge of roads in the image. Inherited from the road map, connectivity and topology are preserved in the process of map conflation. This is advantageous in reconstructing an accurate image road network based on the embedded width data. Derived from the shape of both the curvilinear and junction filter templates, road width can be recovered by making queries to filter stack index maps and retrieving corresponding widths specified as ω in Eq. 3.11 or $\mathbf{w}_+$ in Eq. 4.1. Aligned segments or a single segment are believed to have the consistent width, thus they are enforced to have the same width to avoid the unwanted road width variation. The final output is a binarized *road pixel mask* generated by rasterization of conflated road vectors with an associated width in pixels that is derived during the map conflation process.

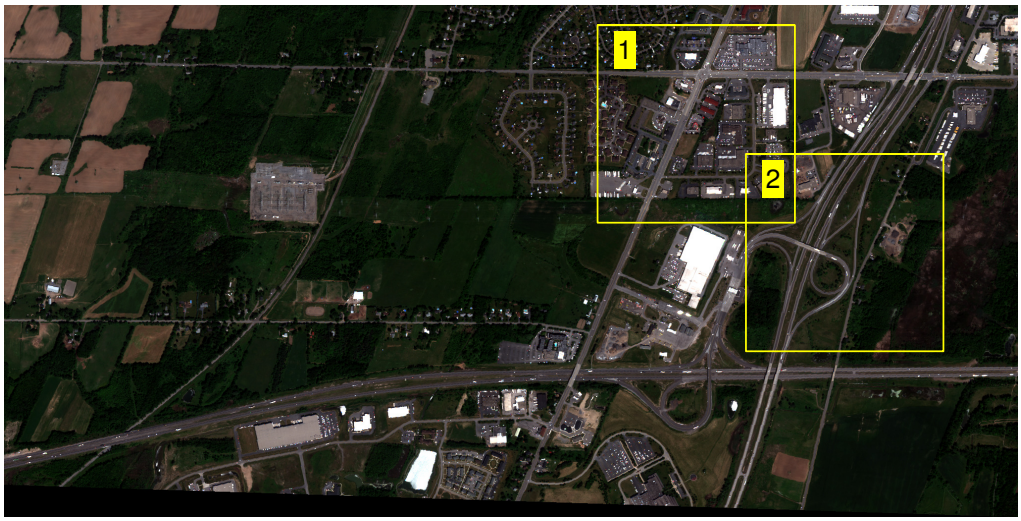


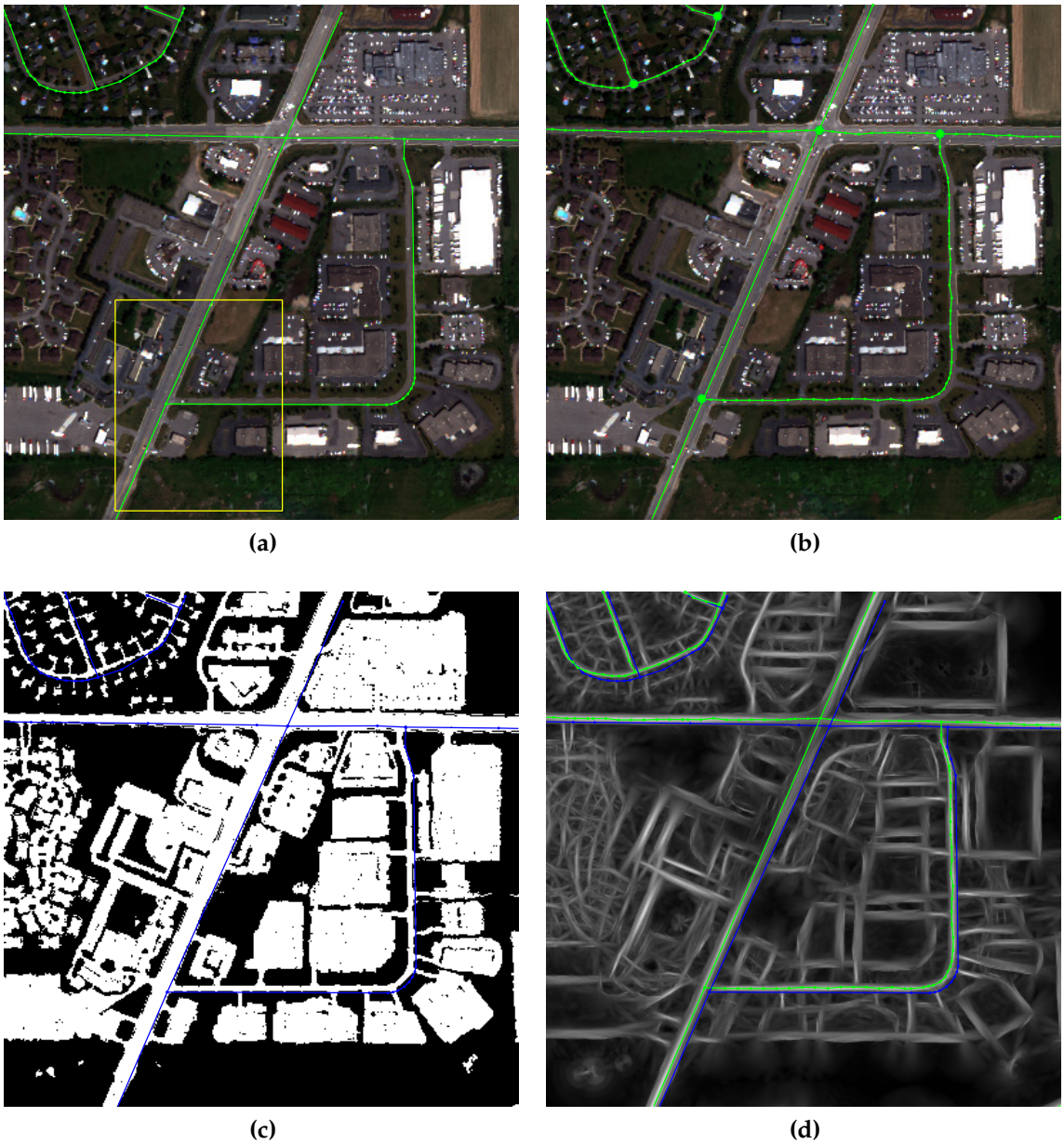
Figure 4.12: *The first image scene shows an suburban area in Henrietta, NY. Two cropped tiles are indicated by yellow boxes.*

4.2 Experimental Results & Discussion

Our approach is tested on four image scenes, whose details can be found in Table 4.1. Images are captured either by WorldView-2 or GeoEye-1 sensors, which are capable of delivering image data with sub-meter resolution and broad spectral bands from the visible to NIR. Two image tiles are cropped from each scene with the size of 400×400 each for multispectral or 1600×1600 each for pan-sharpened. In total there are eight image square tiles, each of which contains scene content that is either typical or challenging for road extraction endeavors. To save the computational time, physical road width is sampled at a few integer values: 8, 12, 16, and 20 (meter), which correspond to the pixel road width ranging from 4 to 10 with the interval of 2 if the GSD is 2 meters.

4.2.1 Image Scene 1

As shown in Fig. 4.12, the first scene covers the suburban area of Henrietta, NY. The number of junctions with valid corrections is 36 out of 43 total junctions.



(a) **(b)** **(c)** **(d)**

Figure 4.13: *1st tile from the 1st scene. (a) RGB image tile with initial road segments (green) overlaid. (b) RGB image tile with conflated road segments (green) overlaid. Solid circles indicate junction points with valid offsets and hollow circles indicate junction points whose offsets are interpolated. (c) BRM overlaid with initial road segments (blue). (d) Maximum projected curvilinear response image overlaid with conflated road segments (green) and initial road segments (blue).*

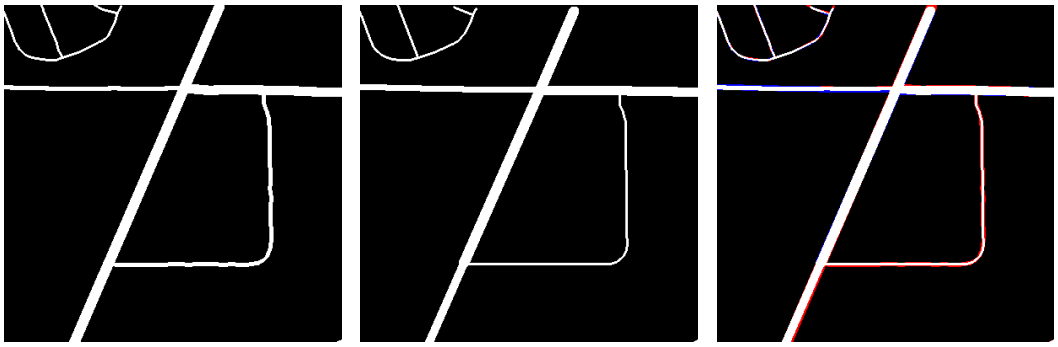


Figure 4.14: *Image road mask of the 1st tile in the 1st scene. (a) shows the extracted road mask. (b) shows the ground truth road mask. (c) shows the comparison of (a) and (b). White pixels represent the pixels that are true positives. Red pixels are false positives. Blue pixels are false negatives.*

Figure 4.15: *RGB image of the 1st tile in the 1st scene with road pixels labeled in magenta.*

The 1st tile displays several typical junctions and road segments with different width as shown in Fig. 4.13a. The initial road network geo-registered to the image deviate from image road centerlines by a varying amount. The BRM of the 1st tile shown in Fig. 4.13c is automatically generated by thresholding the NDVI image ($\tau_{\text{NDVI}} = 0.25$) generated from the multispectral image. In this image tile, a 4-way junction and four 3-way junctions are successfully translated to the image junction points as shown as green solid circles in Fig. 4.13b. The intermediate points also correctly adhere to the corresponding road centerlines. Note that the near-vertical road segments do not have densified intermediate points. In the refinement step these aligned segments and fragments are forced to be collinear with unnecessary points removed. In contrast, other segments are left unchanged with densified points. The road pixels are overlaid with the extracted road mask as shown in Fig. 4.15. Road position and width are captured reasonably well. The road segments connecting the 4-way junction are significantly wider than the other roads. Note that the width of left branch of this 4-way junction is not precisely characterized because the width values within a segment are forced to be consistent to avoid excessive width fluctuation.

The 2nd tile in this scene is shown in Fig. 4.16a. It covers the challenging situation of a highway interchange and its multiple ramps with different curvatures. Road segments are spatially adjacent to or interweaved with each other and the displacement of the roads is relatively large and inconsistent. The matching steps will be briefly reiterated here. First, ramp junction points with valid correction offsets are moved to the updated positions as shown with solid circles in Fig. 4.16b. The hollow circles indicate the junction points whose corrections are invalidated due to the validity checks. Their image positions are assigned by 2-D linear interpolation. Next, intermediate points are tentatively moved based on the positions of validated junction points. Ideally, junction-corrected road segments approach the road centerlines as long as the alignments of junctions are accurate. Vector road segments are satisfactorily conflated to the image road centerlines. Even though there are many curved segments, the curvilinear detector is able to extract their smooth structures because a segment with low-to-medium

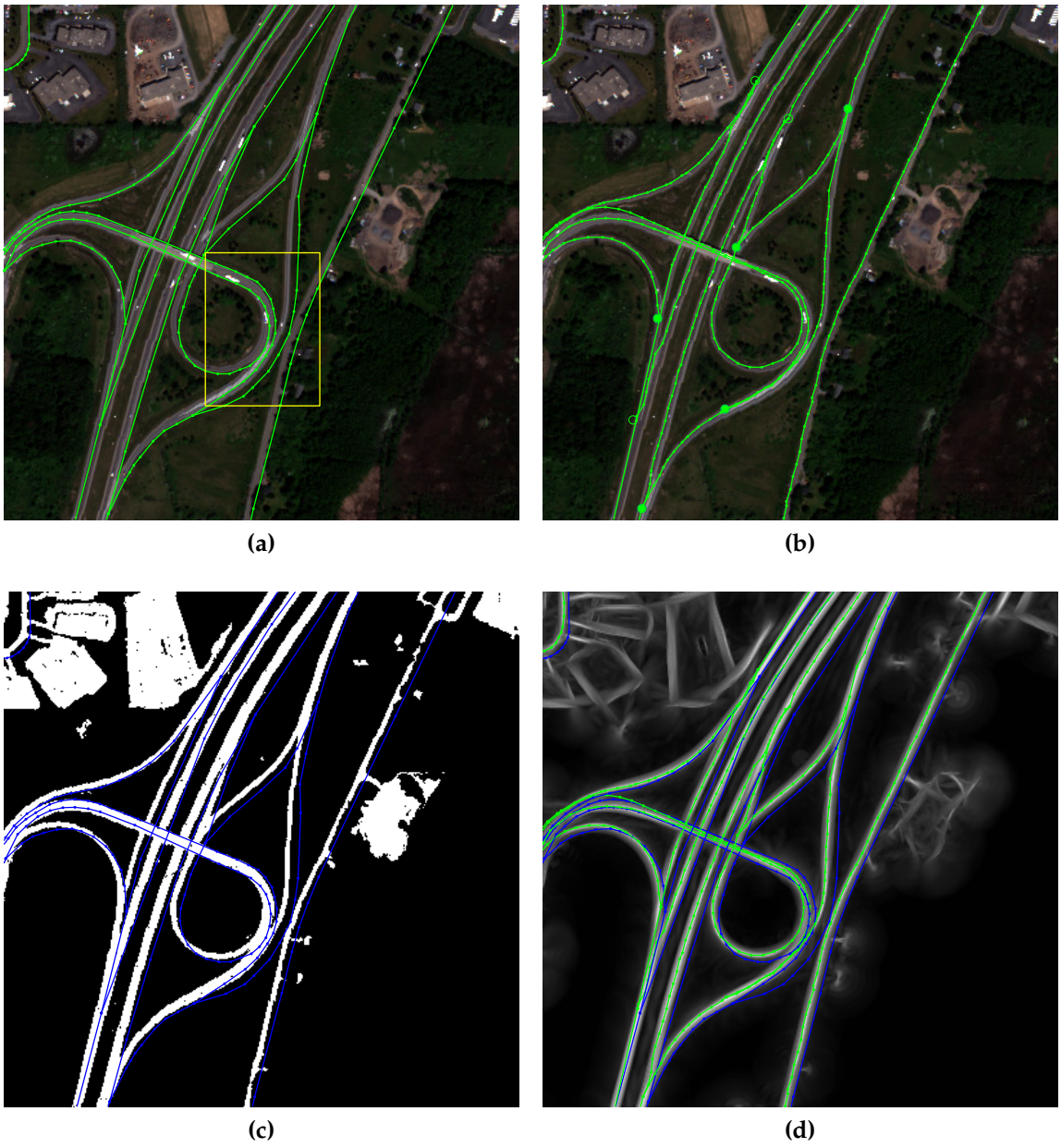
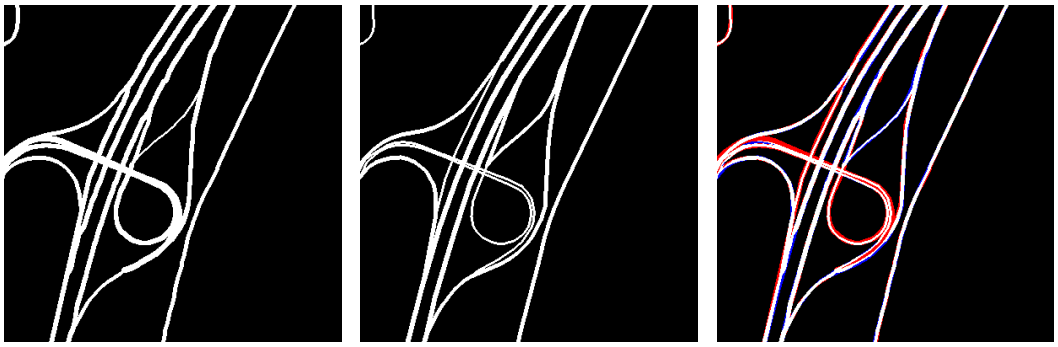


Figure 4.16: *The 2nd tile from the 1st scene. (a) RGB image tile with initial road segments (green) overlaid. (b) RGB image tile with conflated road segments (green) overlaid. Solid circles indicate junction points with valid offsets and hollow circles indicate junction points whose offsets are interpolated. (c) BRM overlaid with initial road segments (blue). (d) Maximum projected curvilinear response image overlaid with conflated road segments (green) and initial road segments (blue).*



(a) (b) (c)

Figure 4.17: *Image road mask of the 2nd tile in 1st scene. (a) shows the extracted road mask. (b) shows the ground truth road mask. (c) shows the comparison of (a) and (b). White pixels represent the pixels that are true positives. Red pixels are those that are false positives. Blue pixels are those that are false negatives.*

Figure 4.18: *RGB image of the 2nd tile in the 1st scene with road pixels labeled in magenta.*

curvature can be effectively approximated as piecewise connected straight lines. The road pixel mask shown in Fig. 4.18 covers the full width of road surface as expected.

4.2.2 Image Scene 2

The second multispectral scene covers a residential area on the east side of Greater Rochester International Airport as shown in Fig. 4.19. Due to the presence of significant tree canopy occlusion, the BRMs (Figs. 4.20c and 4.23c) is created by spectral grouping using ATD, rather than from a binary NDVI, based on manually sampling 7 road pixels. The number of junctions with valid corrections is 68 out of 225 junctions.

The first tile is displayed in Fig. 4.20a. It is chosen to test the robustness of our algorithm when the road surfaces are heavily occluded by tree canopies. The original vector road network is well registered to the image. In this situation, the ideal action would be no action. However, our algorithm cannot be as confident as a human operator, since many road surfaces can be barely seen from above. This poses a big challenge for any existing road extraction approach and can be verified in Fig. 4.20b where most of the junctions are observed to not have valid correction offsets. Only four of them have strong correlation responses in this tile from against either the BRM or the curvilinear response image and are identified as solid circles. These four junctions, along with other junctions with valid corrections that are beyond the tile scope, are used to determine the correction amounts of the rest junctions whose offsets are left unassigned. This strategy preserves the original map topology and suppresses the false positive rate. The design philosophy is that only those junctions with high confidence are corrected and used as control points so that error will not accumulate and propagate along the algorithm pipeline. Similarly, the step of intermediate point matching also finds many points with invalidated corrections after the validity checks. Another difficulty with this tile occurs with the road width estimation. Most junctions and intermediate points do not have valid corrections and the width estimations

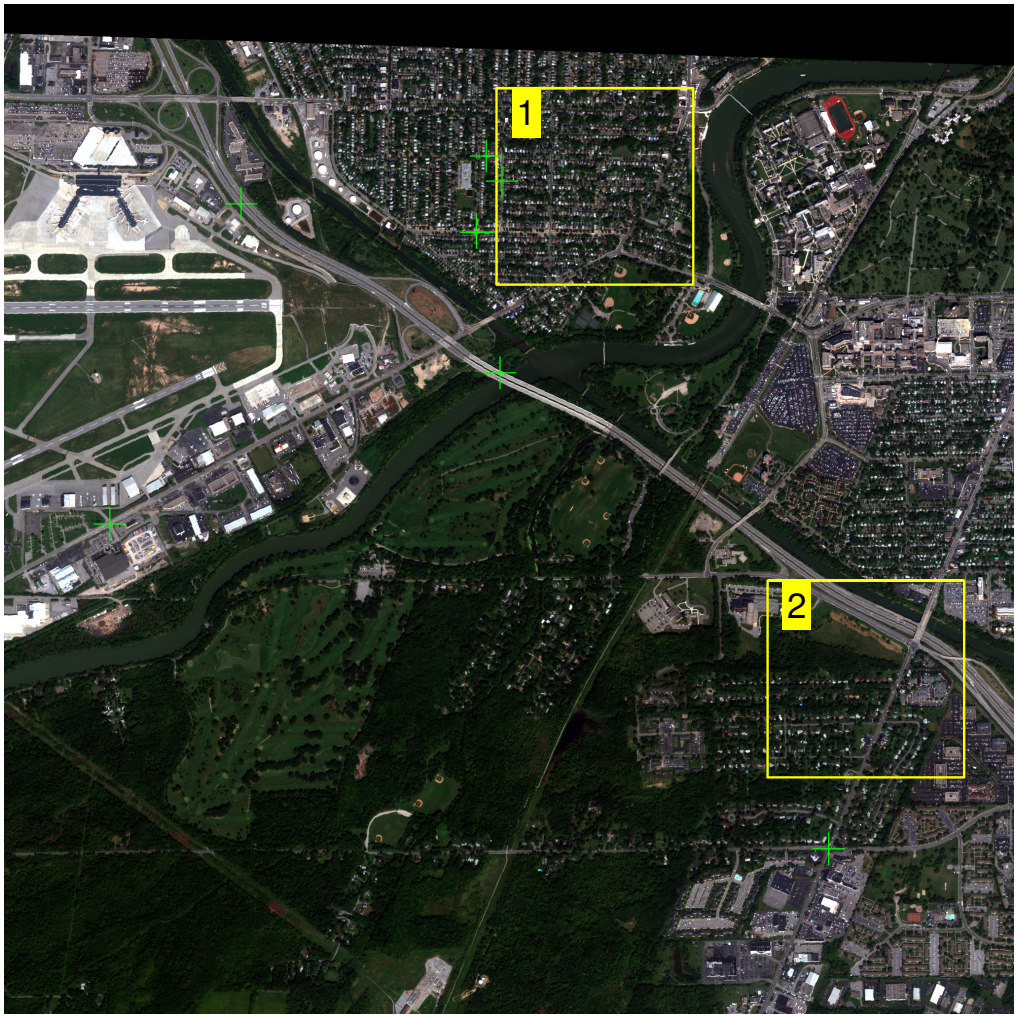


Figure 4.19: *The second image scene shows an area near Greater Rochester International Airport. Two cropped tiles are indicated by yellow boxes. Green pluses represent the selected road sample pixel used with spectral grouping.*

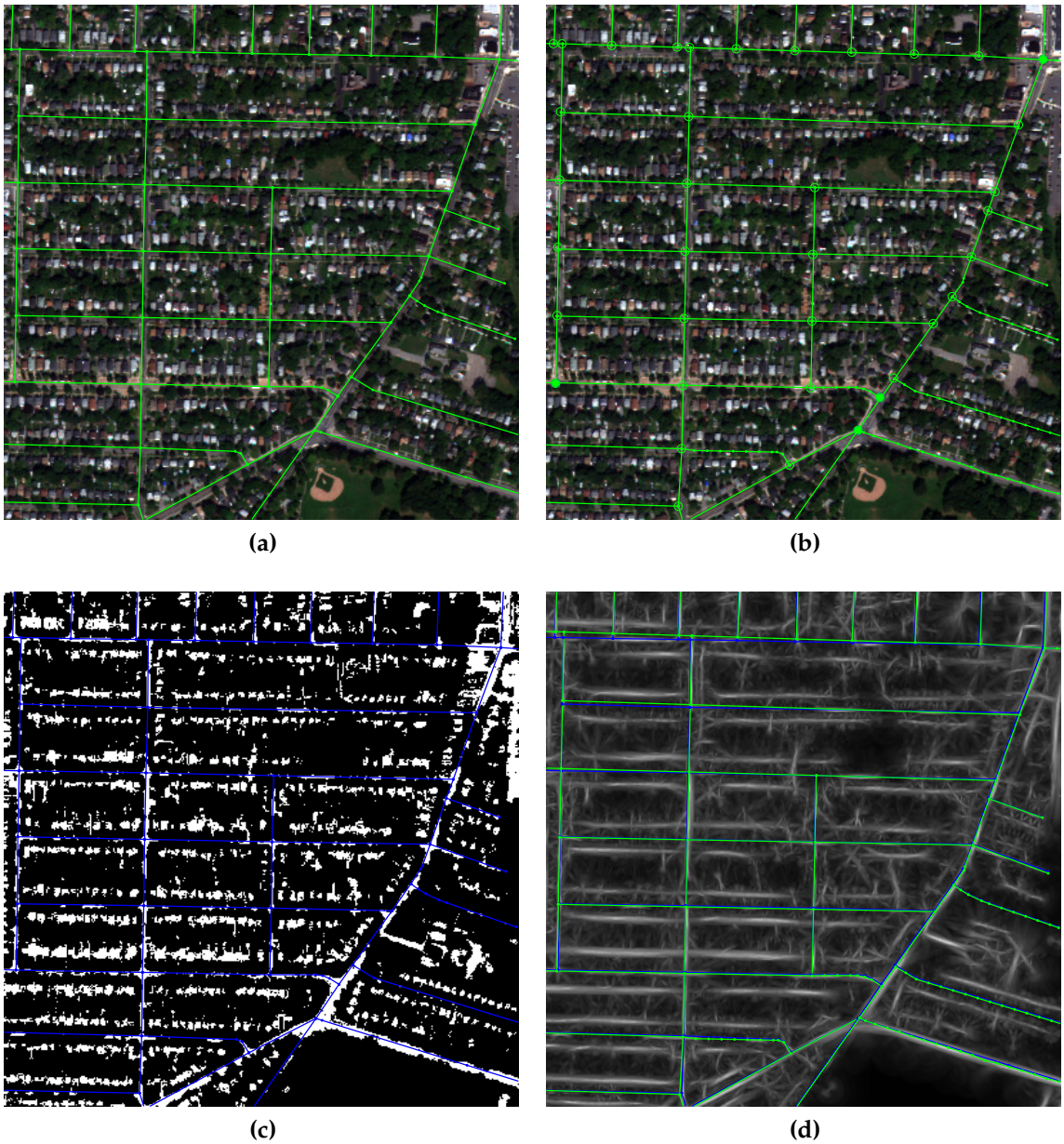


Figure 4.20: *The 1st tile from the 2nd scene. (a) RGB image tile with initial road segments (green) overlaid. (b) RGB image tile with conflated road segments (green) overlaid. Solid circles indicate junction points with valid offsets and hollow circles indicate junction points whose offsets are interpolated. (c) BRM overlaid with initial road segments (blue). (d) Maximum projected curvilinear response image overlaid with conflated road segments (green) and initial road segments (blue).*

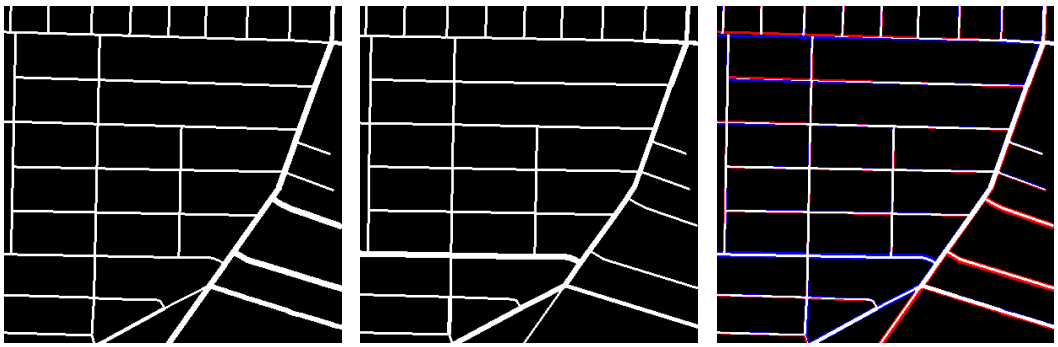


Figure 4.21: *Image road mask of the 1st tile in the 2nd scene. (a) shows the extracted road mask. (b) shows the ground truth road mask. (c) shows the comparison of (a) and (b). White pixels represent the pixels that are true positives. Red pixels are false positives. Blue pixels are false negatives.*

Figure 4.22: *RGB image of the 1st tile in the 2nd scene with road pixels labeled in magenta.*

of the road segments are determined by many fewer points, or none, that pass the validity checks. Hence the width estimates tend to be less accurate given insufficient credible leads.

The second tile displays a similar residential area as the first one with a highway passing by and is shown in Fig. 4.23a. As shown in Fig. 4.23b some junctions (hollow circles) are left unassigned with corrections, which are later interpolated by nearby junctions indicated by solid circles. Though the accuracy of geo-registration is already acceptable, there are a few road segments that require minor adjustments. There is also an incompletely labeled road pavement around the upper-right corner of the image tile shown in Fig. 4.25. The fact that not the entire pavement is labeled is because there is a patch of yellowish lane marker in the middle of that road causing a long hole in binary road map and in turn generates two parallel linear structures. The conflated road segment then deviates from the center and gives an inaccurate width estimation.

4.2.3 Image Scene 3

This image scene focuses on an coastal area in Salvador, Brazil and it has been pan-sharpened to enhance the resolution as shown in Fig. 4.26. This scene poses great challenges for road extraction in an automatic manner. Road pavement materials are diverse and road surfaces are occluded by vehicles, shadows, or buildings. It is highly unlikely that a good extraction outcome can be achieved if no additional road information is exploited, even if supervised learning is attempted. The number of junctions with valid corrections is 213 out of 332. The BRM of the whole scene is again created from spectral grouping using ATD by manually selecting 9 sample pixels. Special attention must be paid in binary mask generation. Many road pixels are cast by shadows and they should be included in the binary set, otherwise these road segments would be impossible to recover in ensuing processing. According to Eq. 3.9, shadow pixels are identified and aggregated to complement the BRM.

The 1st tile shows a dense urban region with little vegetation and many build-

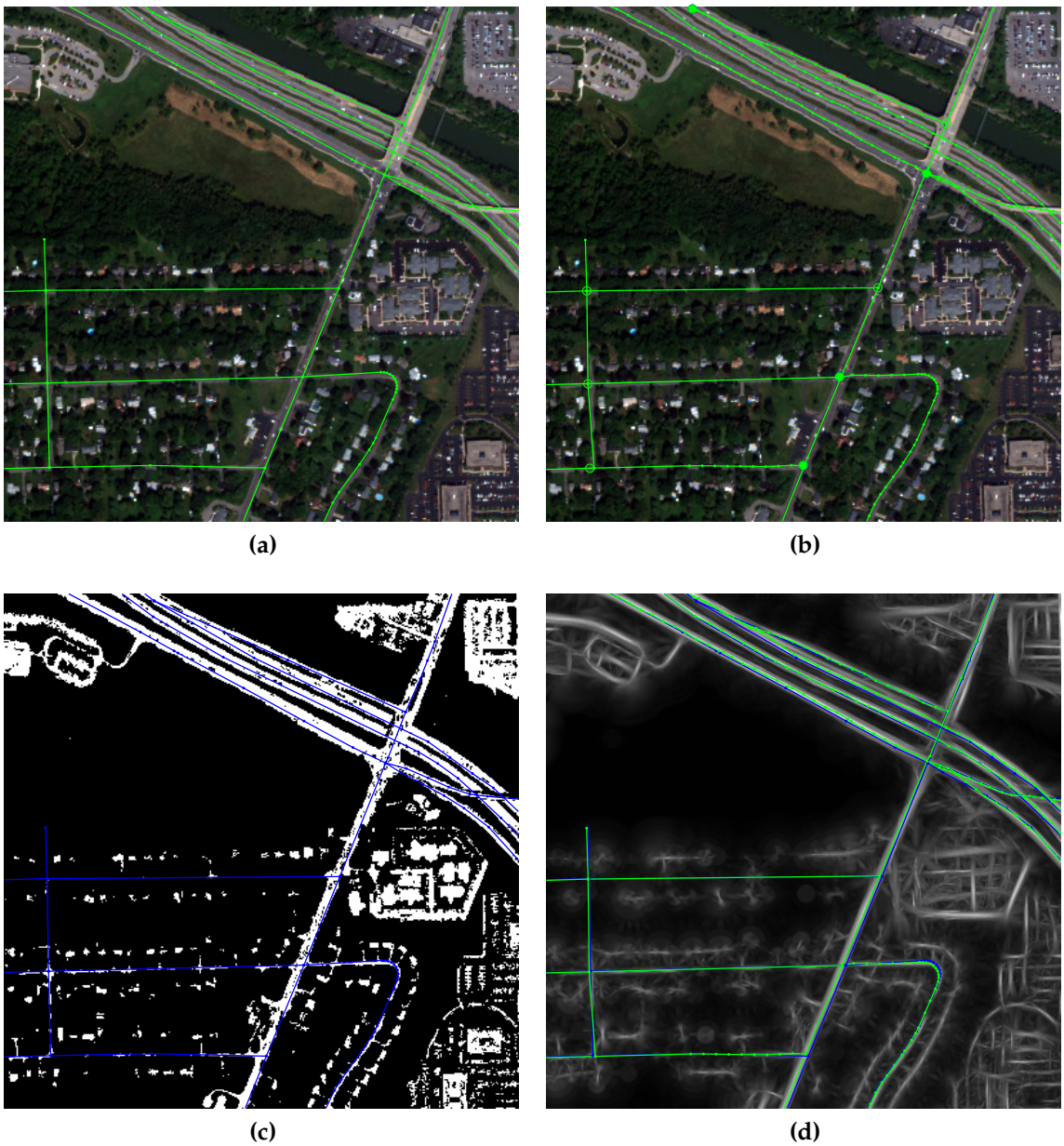


Figure 4.23: *The 2nd tile from the 2nd scene. (a) RGB image tile with initial road segments (green) overlaid. (b) RGB image tile with conflated road segments (green) overlaid. Solid circles indicate junction points with valid offsets and hollow circles indicate junction points whose offsets are interpolated. (c) BRM overlaid with initial road segments (blue). (d) Maximum projected curvilinear response image overlaid with conflated road segments (green) and initial road segments (blue).*

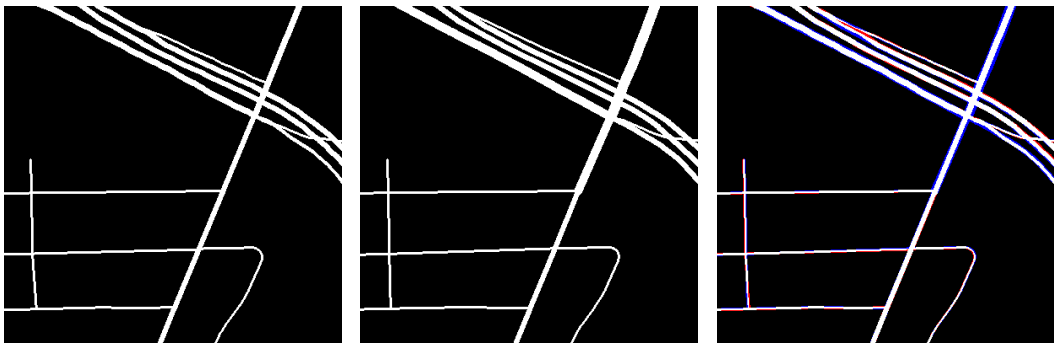


Figure 4.24: *Image road mask of the 2nd tile in the 2nd scene. (a) shows the extracted road mask. (b) shows the ground truth road mask. (c) shows the comparison of (a) and (b). White pixels represent the pixels that are true positives. Red pixels are false positives. Blue pixels are false negatives.*

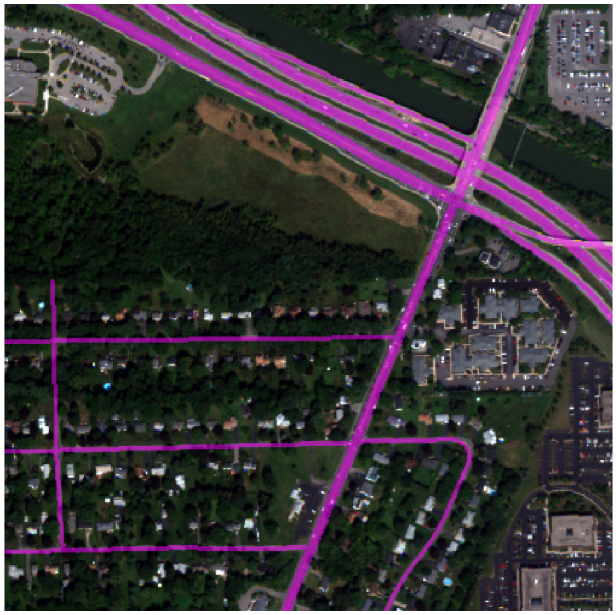


Figure 4.25: *RGB image of the 2nd tile in the 2nd scene with road pixels labeled in magenta.*

Figure 4.26: *The third image scene shows a coastal area in Salvador, Brazil. Two cropped tiles are indicated by yellow boxes. Green pluses represent the nine selected road sample pixels used with spectral grouping.*

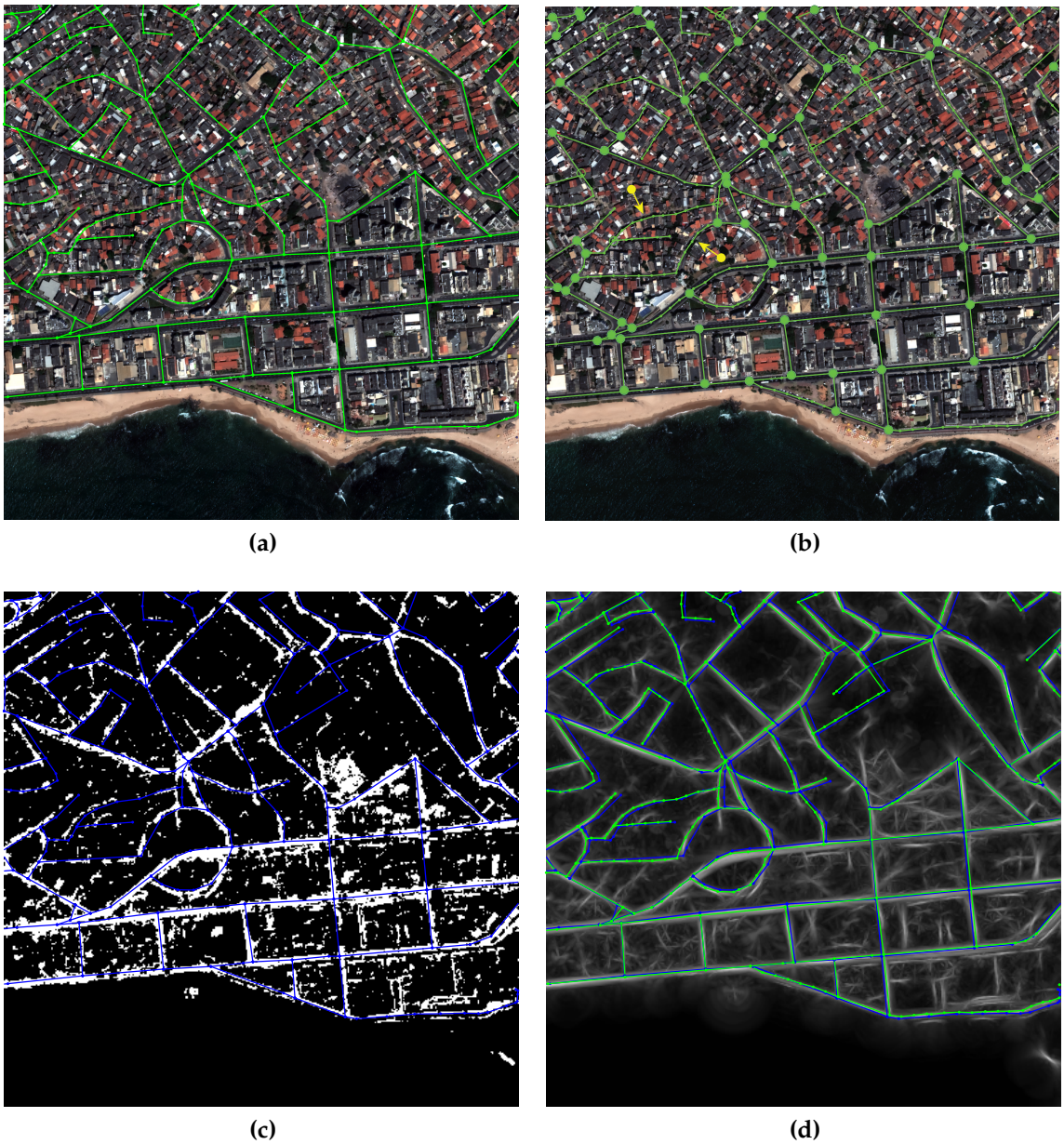


Figure 4.27: *The 1st tile from the 3rd scene. (a) RGB image tile with initial road segments (green) overlaid. (b) RGB image tile with conflated road segments (green) overlaid. Solid circles indicate junction points with valid offsets and hollow circles indicate junction points whose offsets are interpolated. (c) BRM overlaid with initial road segments (blue). (d) Maximum projected curvilinear response image overlaid with conflated road segments (green) and initial road segments (blue).*

Figure 4.28: *Image road mask of 1st tile in the 3rd scene. (a) shows the extracted road mask. (b) shows the ground truth road mask. (c) shows the comparison of (a) and (b). White pixels represent the pixels that are true positives. Red pixels are false positives. Blue pixels are false negatives.*

Figure 4.29: *RGB image of the 1st tile in the 3rd scene with road pixels labeled in magenta.*

ing rooftops spectrally similar to the roads. The original road network is close to, but not precisely aligned with the road centerlines as shown in Fig. 4.27a. There are many undetermined junction points (hollow circles) highlighted in Fig. 4.27b, which confirms the scene complexity. However, the method is successful and conflated roads lie exactly on the road centerlines. Some noticeable corrections are made to the road segments indicated by the arrows. The road segment indicated by one arrow fits well in the shadowed road and that indicated by another arrow becomes rounder and matches better with the corresponding road features. The extracted road pixel mask is compared with the ground truth as shown in Fig. 4.28. Also referring to Fig. 4.29, the width accuracy is fairly good for most road segments.

The 2nd tile displays a less densely developed urban site, with many tall buildings casting shadows on the roads. The method makes changes to many of the original road segments. Some corrections made to the segments are indicated by the arrows in Fig.4.30b. Junction in shadow can be located by means of interpolation. The shadowed road segments are also entirely recovered and the road connectivity is preserved. The extracted road pixel mask again matches well with the image roads as shown in Fig. 4.32.

4.2.4 Image Scene 4

The fourth image scene covers a large geographical area of Rome, Italy as shown in Fig. 4.33. Two image tiles are selected to demonstrate the performances of our algorithm on a dense urban area (the 1st tile) and a hilly rural area (the 2nd tile).

Spectral grouping using ATD is the way to generate the BRMs for both tiles. The 1st tile presents a very busy urban scene (Fig. 4.34) and the overlaid road vectors are obviously misaligned with the image. Again conflation works as desired and as shown in Fig. 4.36 the extracted road network is quite complete and accurate. Referring to Fig. 4.37, the corrections imposed on the 2nd tile are relatively minor and curved road segments are conflated successfully. The final road pixel mask is shown in Fig. 4.39.

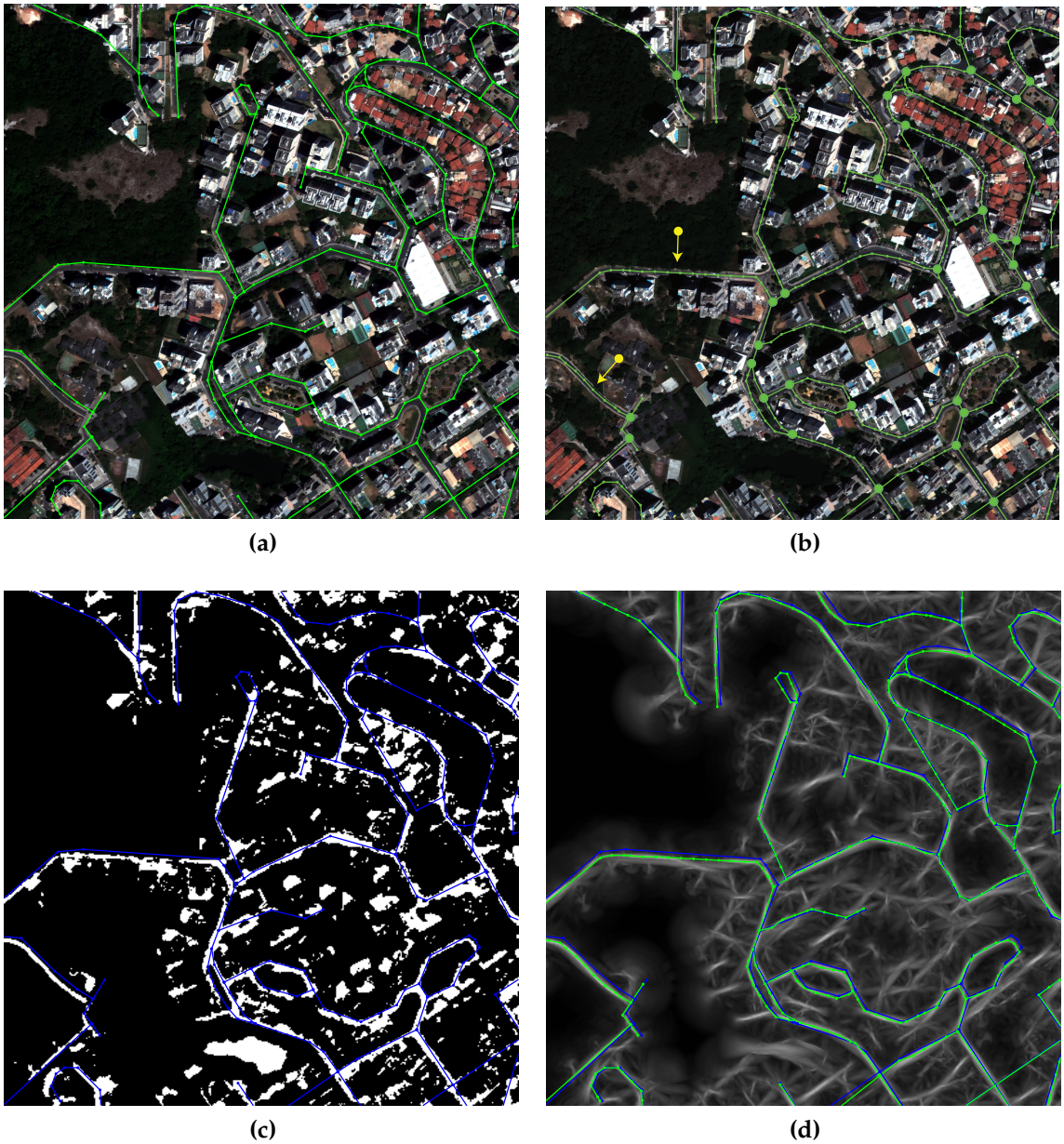


Figure 4.30: *The 2nd tile from the 3rd scene. (a) RGB image tile with initial road segments (green) overlaid. (b) RGB image tile with conflated road segments (green) overlaid. Solid circles indicate junction points with valid offsets and hollow circles indicate junction points whose offsets are interpolated. (c) BRM overlaid with initial road segments (blue). (d) Maximum projected curvilinear response image overlaid with conflated road segments (green) and initial road segments (blue).*

Figure 4.31: *Image road mask of the 2nd tile in the 3rd scene. (a) shows the extracted road mask. (b) shows the ground truth road mask. (c) shows the comparison of (a) and (b). White pixels represent the pixels that are true positives. Red pixels are false positives. Blue pixels are false negatives.*

Figure 4.32: *RGB image of the 2nd tile in the 3rd scene with road pixels labeled in magenta.*

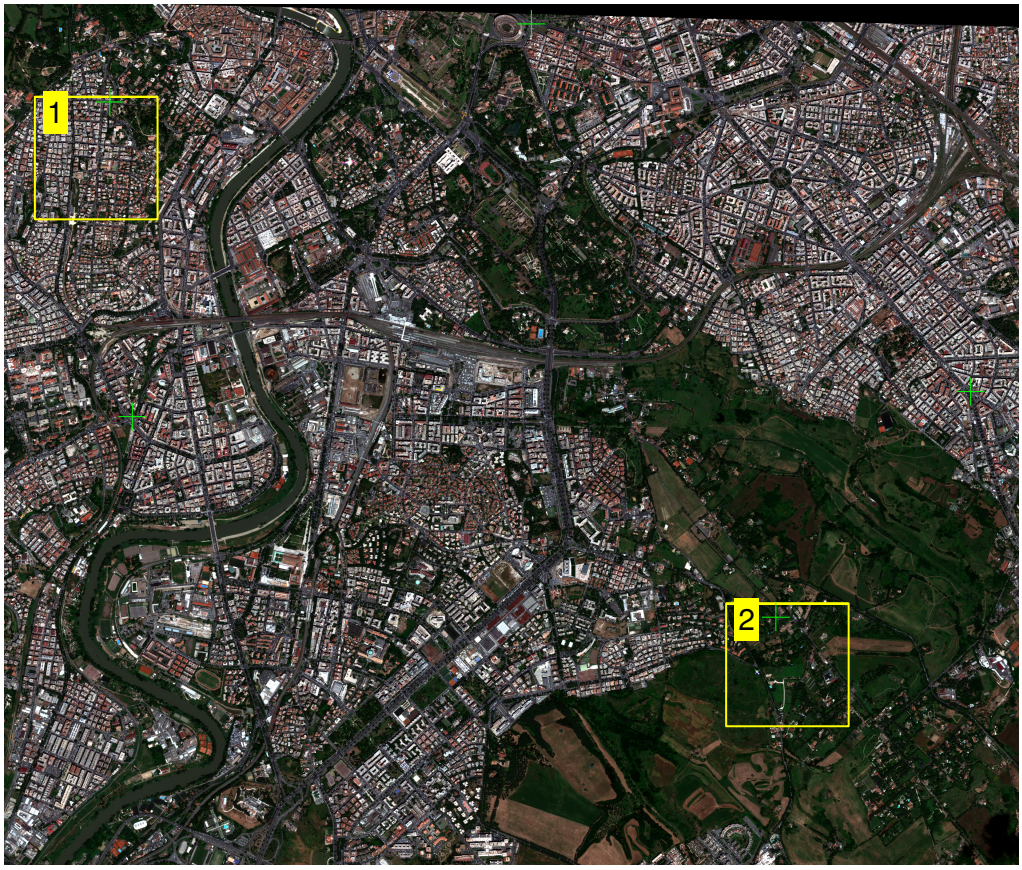


Figure 4.33: *The fourth image scene shows a combined urban and rural area near Rome, Italy. Two cropped tiles are indicated by yellow boxes. Green pluses represent the five selected road sample pixels used with spectral grouping.*

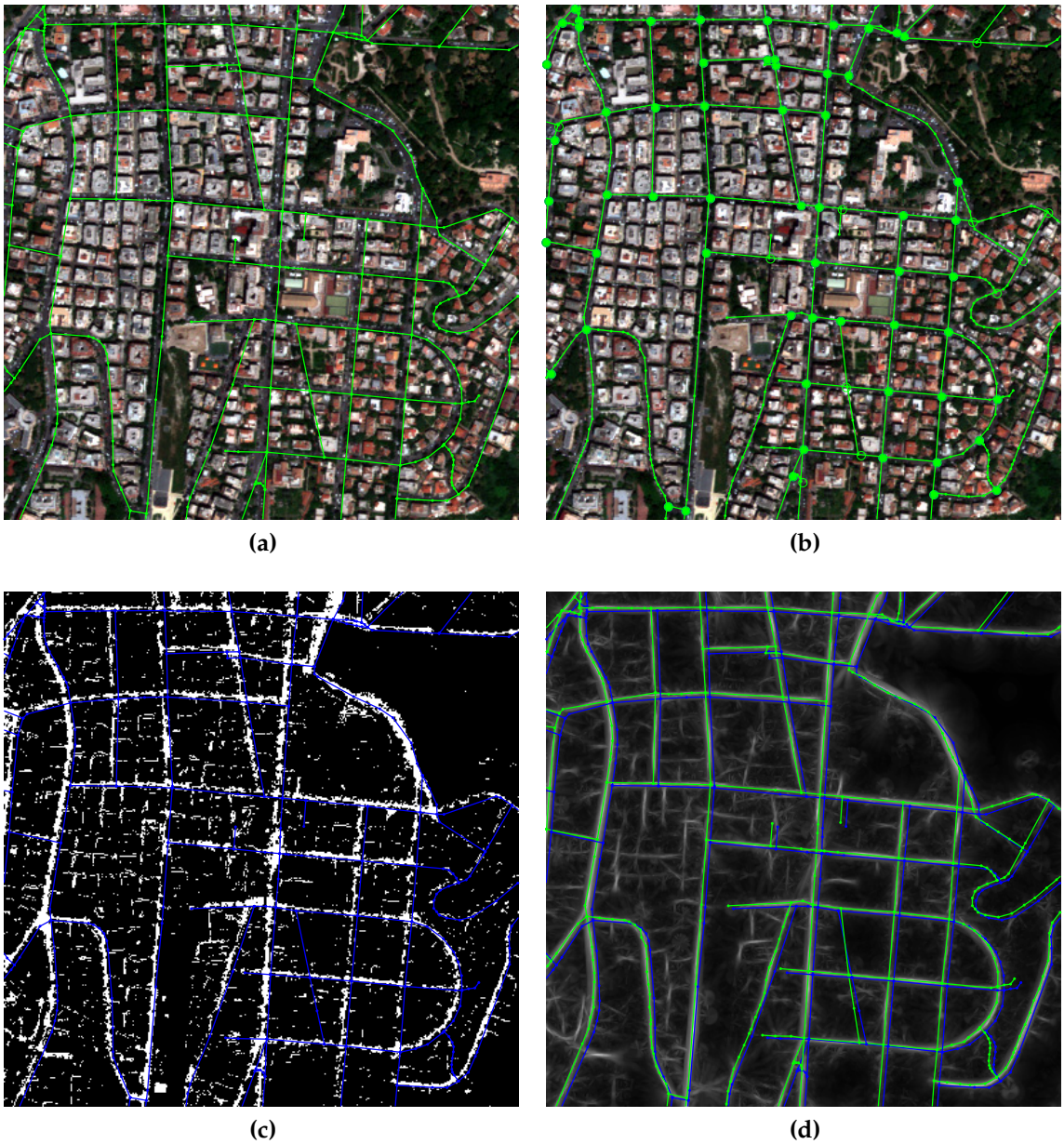


Figure 4.34: *The 1st tile from the 4th image scene. (a) RGB image tile with initial road segments (green) overlaid. (b) RGB image tile with conflated road segments (green) overlaid. Solid circles indicate junction points with valid offsets and hollow circles indicate junction points whose offsets are interpolated. (c) BRM overlaid with initial road segments (blue). (d) Maximum projected curvilinear response image overlaid with conflated road segments (green) and initial road segments (blue).*

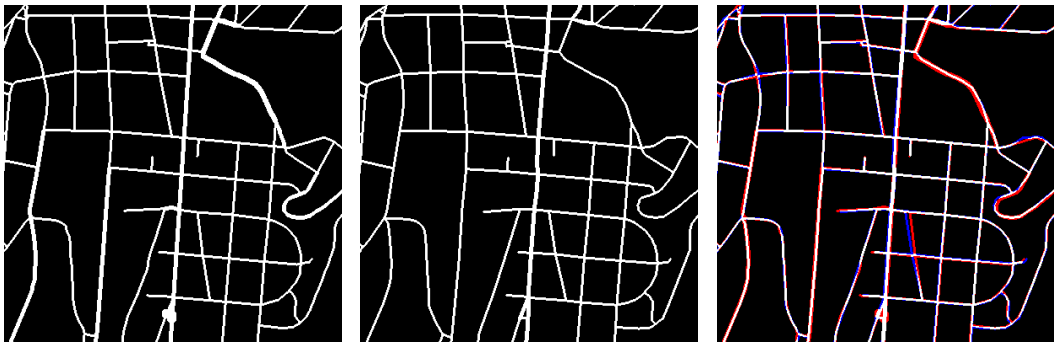


Figure 4.35: *Image road mask of the 1st tile in the 4th scene. (a) shows the extracted road mask. (b) shows the ground truth road mask. (c) shows the comparison of (a) and (b). White pixels represent the pixels that are true positives. Red pixels are false positives. Blue pixels are false negatives.*

Figure 4.36: *RGB image of the 1st tile in the 4th scene with road pixels labeled in magenta.*

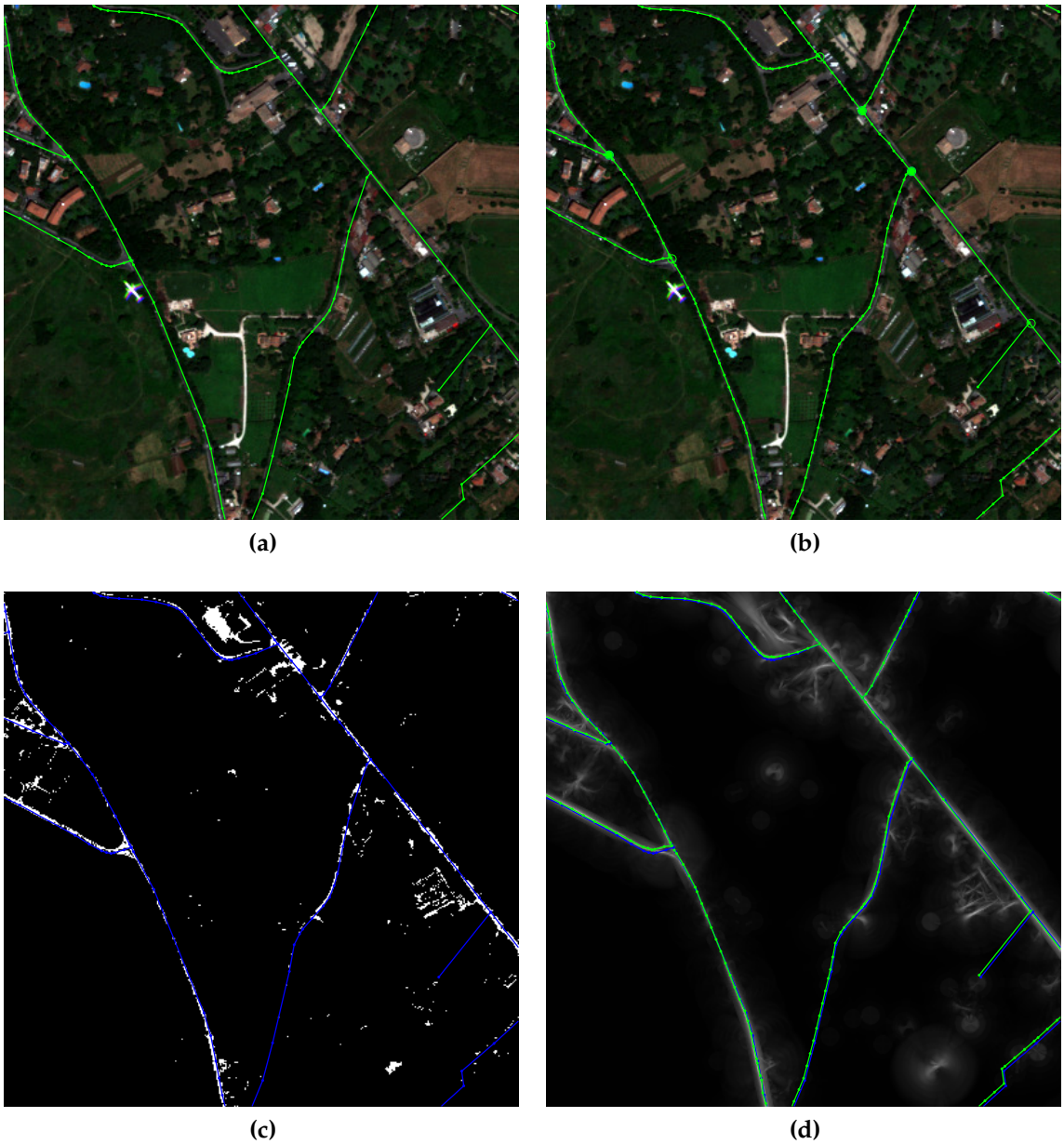


Figure 4.37: *The 2nd tile from the 4th image scene. (a) RGB image tile with initial road segments (green) overlaid. (b) RGB image tile with conflated road segments (green) overlaid. Solid circles indicate junction points with valid offsets and hollow circles indicate junction points whose offsets are interpolated. (c) BRM overlaid with initial road segments (blue). (d) Maximum projected curvilinear response image overlaid with conflated road segments (green) and initial road segments (blue).*

Figure 4.38: *Image road mask of the 2nd tile in the 4th scene. (a) shows the extracted road mask. (b) shows the ground truth road mask. (c) shows the comparison of (a) and (b). White pixels represent the pixels that are true positives. Red pixels are false positives. Blue pixels are false negatives.*

Figure 4.39: *RGB image of the 2nd tile in the 4th scene with road pixels labeled in magenta.*

4.2.5 Discussion

Large and diverse image data sets are tested using our proposed algorithm pipeline. To quantitatively evaluate the overall performance of the system workflow, two common measures are applied to assess the accuracies of both image road extraction and vector road conflation. Precision and recall are used to evaluate the system performance and are defined as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4.4}$$

and

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4.5}$$

respectively, where TP represents true positive, FP represents false positive, and FN represents false negative.

Ground truth of image road pixels is generated by first extracting the road skeletons using QGIS tools and then expanding the skeletons with the manually determined width for each segment. Based on the ground truth, both precision and recall are computed to assess the detection performance. Moreover, 3-pixel (full width) buffer zone is created based on ground truth vector road so as to assess the conflation correctness. The former measure is a more rigorous evaluation criterion since no buffer zone is allowed in comparison. The ground truth road pixel masks and the extracted road masks for all eight image tiles in four scenes are shown for performance comparisons in Figs. 4.14, 4.17, 4.21, 4.24, 4.28, 4.31, 4.35, and 4.38, respectively.

The proposed approach works well on several challenging scenes and has been shown to generate road extraction accuracy as high as 91.32% and 92.58% for precision and recall, respectively. Conflation correctness is up to 94.93%. The accuracy statistics are summarized in Table 4.2. Though the recall percentages of the 2nd scene are below 80%, the high conflation accuracy still holds. Likewise, for the 3rd and 4th scenes, which are very challenging for road extraction

Scene-Tile	Road Extraction		Conflation
	Precision	Recall	Correctness
1-1	86.92%	92.58%	94.93%
1-2	78.49%	87.51%	80.37%
2-1	72.66%	73.21%	87.21%
2-2	91.32%	78.57%	92.84%
3-1	78.87%	83.54%	89.70%
3-2	72.42%	78.96%	88.24%
4-1	72.57%	80.94%	88.21%
4-2	76.46%	77.21%	85.45%

Table 4.2: *Accuracy statistics on the test image scenes.*

algorithms, the road extraction accuracies range from 72% to 84%. The conflation correctness, however, is still above 85% for both scenes, which clearly justifies the effectiveness of our algorithm pipeline. In addition, to the best of our knowledge there is no published approach that has been ever attempted on any scenes as complex as the 3rd scene. Therefore the performance of our extraction system appears to support its robustness.

In this section, we quantitatively tested our approach on eight representative image tiles encountered in urban and rural road extraction. Though the performance is quite satisfactory in the above demonstration, a few algorithm issues require further investigation. For example, road width or orientation estimation is sometimes not very accurate because the roadside objects may obscure some portions of the lanes, in which case the extracted road features may not be able to convey the real road characteristics. Double parallel roads close to each other tend to merge into a single lane after conflation. Imposing minimum lateral distance on conflation of parallel roads could ensure that they remain separate.

Our algorithm only requires localized operations in a neighborhood region and thus can be easily parallelized on partitioned image tiles to boost computation. Its scalability assures its possible extension to work efficiently on very large scale images. Our system can also be readily adapted to imagery-to-vector conflation by using pairs of image junctions and map junctions with valid alignments

as control point pairs and applying thin-plate-spline transform on the image.

4.3 Summary

The road extraction stage of the system workflow is presented in this chapter. Our approach includes both image-based road extraction and road map conflation analysis in a combined multi-step process. It utilizes road features extracted from the previous stage to automatically conflate a vector road map to a geo-referenced multispectral image and simultaneously extract road pixels from that image.

The proposed approach for road extraction using a road map is generic since OpenStreetMap vector road data is globally and readily available and also free of charge and usage restriction. More importantly, its map quality improves and coverage grows over time, which makes it valuable prior data for image-based road extraction. Due to the persistent mis-registration between image and map data, map conflation is carried to align road vectors with road centerlines in the image. Junction templates derived from rasterized road segments are matched with both an image-derived binary road mask and a curvilinear response image to conflate junction points. Next, the intermediate point matching step effectively conflates the vector road network based on junction-corrected road segments and image curvilinear structures, within which width and orientation estimations are also embedded and can be obtained to recover the piecewise width of road segments. A road pixel mask is finally created with complete road knowledge - centerline, width, connectivity, and topology. Our approach was tested on some large and diverse image data sets and was verified for its effectiveness and robustness in achieving a minimum of 80% conflation correctness and 70% road extraction accuracy on some extremely challenging scenes that have not been attempted in prior work.

Chapter 5

Conclusions and Future Work

The objective of this research is to: 1) exploit representative road features to confidently determine the presence and characteristics of a road network in an image; 2) based on the extracted road features, develop a vector road map to raster imagery conflation algorithm; 3) based on the automatically conflated map, implement the image-based road network extraction.

Chapter 2 presented a novel spectral similarity measure, ATD (anisotropy-tunable distance), that can be used to differentiate spectral changes in both magnitude and direction in a user adjustable manner. Also a simple radiometric calibration approach was proposed to reinforce data collinearity as a pre-processing step.

Chapter 3 provided the details of the feature extraction stage of the algorithm pipeline. NNDiffuse (nearest neighbor diffusion), an efficient pan-sharpening algorithm, was utilized as an optional step to create a resolution-enhanced multispectral image. Either Binarized NDVI or spectral grouping using ATD was then used to create a BRM (binary road mask). A curvilinear structure response image was also generated to provide unique linear road features. Finally, vector and rasterized road maps were created as the sources of geometric road features.

Chapter 4 elaborated the road extraction stage of the algorithm pipeline. Road features were used to conflate road junctions and intermediate points. The con-

flated road network aided in road pixel extraction from a multispectral image. The whole system was tested on a set of challenging, large-scale, and diverse image scenes and its performance was verified to be fairly accurate and robust.

Our proposed approach is capable of conflating a misaligned road map to a geo-referenced multispectral image and extracting road pixels from that image in an integral system. The main contributions of our work are listed below:

- A system workflow that seamless integrates the tasks of image-based road network extraction and automated vector road to raster imagery conflation.
- A image processing algorithm pipeline dedicated to extract complete knowledge of the road network from a geo-referenced multispectral image, which involves spectral signature, geographic location, lane width, orientation, topology, and even possible occlusion.
- A set of informative and robust spectral/spatial road features engineered to apply hierarchically on complex image scenes.
- A novel spectral similarity measure that flexibly accounts for both changes of spectral magnitude and direction and is suited for spectral grouping in road pixel detection in a multispectral image.

Several points are provided here for future research directions:

- Improve the performance of road extraction in complex scenes by introducing more relevant knowledge about road characteristics.
- Optimize junction matching method when the shape of junction vector differs significantly from the image junction.
- Further study could focus on adaptive selection of ATD parameters based on scene contents.

Bibliography

[1] D. Chai, W. Forstner, and F. Lafarge, "Recovering line-networks in images by junction-point processes," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 1894–1901.

[2] J. Schott, *Remote Sensing: The Image Chain Approach*, New York: Oxford University Press, 2nd ed., 2007.

[3] W. Sun, B. Chen, and D. W. Messinger, "Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images," *Optical Engineering* **53**(1), pp. 013107–013107, 2014.

[4] J. B. Mena, "State of the art on automatic road extraction for GIS update: a novel classification," *Pattern Recognition Letters* **24**(16), pp. 3037–3058, 2003.

[5] Q. Zhang, J. Wang, X. Peng, P. Gong, and P. Shi, "Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data," *International Journal of Remote Sensing* **23**(15), pp. 3057–3078, 2002.

[6] G. Conte and P. Doherty, "An integrated uav navigation system based on aerial image matching," in *Aerospace Conference, 2008 IEEE*, pp. 1–10.

[7] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Robust inference of principal road paths for intelligent transportation systems," *Intelligent Transportation Systems, IEEE Transactions on* **12**(1), pp. 298–308, 2011.

[8] R. Bajcsy and M. Tavakoli, "Computer recognition of roads from satellite pictures," *Systems, Man and Cybernetics, IEEE Transactions on* **SMC-6**(9), pp. 623–637, 1976.

[9] J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka, "Road network extraction and intersection detection from aerial images by tracking road footprints," *Geoscience and Remote Sensing, IEEE Transactions on* **45**(12), pp. 4144–4157, 2007.

[10] X. Hu and V. Tao, "Automatic extraction of main road centerlines from high resolution satellite imagery using hierarchical grouping," *Photogrammetric Engineering and Remote Sensing* **73**(9), p. 1049, 2007.

[11] S. Das, T. T. Mirnalinee, and K. Varghese, "Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images," *Geoscience and Remote Sensing, IEEE Transactions on* **49**(10), pp. 3906–3931, 2011.

[12] X. Lin, J. Zhang, Z. Liu, J. Shen, and M. Duan, "Semi-automatic extraction of road networks by least squares interlaced template matching in urban areas," *International Journal of Remote Sensing* **32**(17), pp. 4943–4959, 2011.

[13] C. Unsalan and B. Sirmacek, "Road network detection using probabilistic and graph theoretical methods," *Geoscience and Remote Sensing, IEEE Transactions on* **50**(11), pp. 4441–4453, 2012.

[14] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler, "A higher-order crf model for road network extraction," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 1698–1705.

[15] W. Sun and D. W. Messinger, "Knowledge-based automated road network extraction system using multispectral images," *Optical Engineering* **52**(4), pp. 047203–047203, 2013.

[16] W. Shi, Z. Miao, and J. Debayle, "An integrated method for urban main-road centerline extraction from optical remotely sensed imagery," *Geoscience and Remote Sensing, IEEE Transactions on* **PP**(99), pp. 1–14, 2013.

[17] C. Zhang, "Towards an operational system for automated updating of road databases by integration of imagery and geodata," *ISPRS Journal of Photogrammetry and Remote Sensing* **58**(3–4), pp. 166–186, 2004.

[18] X. Hu, C. V. Tao, and Y. Hu, "Automatic road extraction from dense urban area by integrated processing of high resolution imagery and lidar data," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Istanbul, Turkey* **35**, p. B3, 2004.

[19] C. Poullis and S. You, "Delineation and geometric modeling of road networks," *ISPRS Journal of Photogrammetry and Remote Sensing* **65**(2), pp. 165–181, 2010.

[20] J. Zhao and Y. Suya, "Road network extraction from airborne lidar data using scene context," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pp. 9–16.

[21] W. B. Song, J. M. Keller, T. L. Haithcoat, and C. H. Davis, "Automated geospatial conflation of vector road maps to high resolution imagery," *Image Processing, IEEE Transactions on* **18**(2), pp. 388–400, 2009.

[22] E. P. Baltsavias, "Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems," *ISPRS Journal of Photogrammetry and Remote Sensing* **58**(3–4), pp. 129–151, 2004.

[23] C.-C. Chen, C. A. Knoblock, and C. Shahabi, "Automatically conflating road vector data with orthoimagery," *GeoInformatica* **10**(4), pp. 495–530, 2006.

[24] X. Wu, R. Carceroni, H. Fang, S. Zelinka, and A. Kirmse, "Automatic alignment of large-scale aerial rasters to road-maps," in *Proceedings of the 15th*

Annual ACM International Symposium on Advances in Geographic Information Systems, *GIS '07*, pp. 17:1–17:8, ACM, (New York, NY, USA), 2007.

[25] T. Peng, I. H. Jermyn, V. Prinet, and J. Zerubia, "Incorporating generic and specific prior knowledge in a multiscale phase field model for road extraction from VHR images," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **1**(2), pp. 139–146, 2008.

[26] W. Song, *Automated vector-vector and vector-imagery geospatial conflation*, University of Missouri at Columbia, 2011.

[27] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *Computer Vision–ECCV 2010*, pp. 210–223, Springer, 2010.

[28] J. Zhang, Y. Zhu, and L. Meng, "Conflation of road network and georeferenced image using sparse matching," in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 281–288, ACM, 2011.

[29] L. Lu, Y. Zhang, P. Tao, Z. Zhang, and Y. Zhang, "Estimation of transformation parameters between centre-line vector road maps and high resolution satellite images," *The Photogrammetric Record* **28**(142), pp. 130–144, 2013.

[30] J. Yuan and A. M. Cheriyadat, "Road segmentation in aerial images by exploiting road vector data," in *Computing for Geospatial Research and Application (COM.Geo), 2013 Fourth International Conference on*, pp. 16–23.

[31] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, "The spectral image-processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment* **44**(2-3), pp. 145–163, 1993.

[32] C. I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *Information Theory, IEEE Transactions on* **46**(5), pp. 1927–1932, 2000.

[33] C. I. Chang, *Hyperspectral imaging: Techniques for spectral detection and classification*, vol. 1, Springer US, 2003.

[34] F. van der Meer and W. Bakker, "Cross correlogram spectral matching: Application to surface mineralogical mapping by using AVIRIS data from Cuprite, Nevada," *Remote Sensing of Environment* **61**(3), pp. 371–382, 1997.

[35] O. A. Carvalho Junior and P. R. Meneses, *Spectral correlation mapper (SCM): An improvement on the spectral angle mapper (SAM)*. Proc. 9th Airborne Earth Sci. Workshop, JPL Pub. 00-18, 2000.

[36] P. Mahalanobis, "On the generalized distance in statistics," in *Proceedings of the National Institute of Sciences of India*, **2**(1), pp. 49–55, New Delhi, 1936.

[37] F. van der Meer, "The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery," *International Journal of Applied Earth Observation and Geoinformation* **8**(1), pp. 3–17, 2006.

[38] J. Farifteh, F. van der Meer, and E. J. M. Carranza, "Similarity measures for spectral discrimination of salt-affected soils," *International Journal of Remote Sensing* **28**(23), pp. 5273–5293, 2007.

[39] B. D. Bue, E. Merenyi, and B. Csatho, "Automated labeling of materials in hyperspectral imagery," *Geoscience and Remote Sensing, IEEE Transactions on* **48**(11), pp. 4059–4070, 2010.

[40] M. W. Matthew, S. M. Adler-Golden, A. Berk, G. Felde, G. P. Anderson, D. Gorodetzky, S. Paswaters, and M. Shippert, "Atmospheric correction of spectral imagery: evaluation of the FLAASH algorithm with AVIRIS data," in *Applied Imagery Pattern Recognition Workshop, 2002. Proceedings. 31st*, pp. 157–163, IEEE, 2002.

[41] L. S. Bernstein, S. M. Adler-Golden, R. L. Sundberg, R. Y. Levine, T. C. Perkins, A. Berk, A. J. Ratkowski, G. Felde, and M. L. Hoke, "A new method

for atmospheric correction and aerosol optical property retrieval for VIS-SWIR multi-and hyperspectral imaging sensors: QUAC (quick atmospheric correction)," tech. rep., DTIC Document, 2005.

[42] R. Wolfe, J. Masek, N. Saleous, and F. Hall, "LEDAPS: mapping North American disturbance from the Landsat record," in *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International*, **1**, IEEE, 2004.

[43] P. Keranen, A. Kaarna, and P. J. Toivanen, "Spectral similarity measures for classification in lossy compression of hyperspectral images," in *International Symposium on Remote Sensing*, pp. 285–296, International Society for Optics and Photonics, 2003.

[44] Y. Z. Du, C. I. Chang, H. Ren, C. C. Chang, J. O. Jensen, and F. M. D'Amico, "New hyperspectral discrimination measure for spectral characterization," *Optical Engineering* **43**(8), pp. 1777–1786, 2004.

[45] M. N. Kumar, M. V. R. Seshasai, K. S. V. Prasad, V. Kamala, K. V. Ramana, R. S. Dwivedi, and P. S. Roy, "A new hybrid spectral similarity measure for discrimination among vigna species," *International Journal of Remote Sensing* **32**(14), pp. 4041–4053, 2011.

[46] O. A. Carvalho Junior, R. F. Guimaraes, A. R. Gillespie, N. C. Silva, and R. A. T. Gomes, "A new approach to change vector analysis using distance and similarity measures," *Remote Sensing* **3**(11), pp. 2473–2493, 2011.

[47] A. Galal, H. Hassan, and I. F. Imam, "A novel approach for measuring hyperspectral similarity," *Applied Soft Computing* **12**(10), pp. 3115–3123, 2012.

[48] R. R. Nidamanuri and B. Zbell, "Normalized spectral similarity score (NS3) as an efficient spectral library searching method for hyperspectral image classification," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **4**(1), pp. 226–240, 2011.

[49] B. D. Bue and E. Merenyi, "An adpative similarity measure for classification of hyperspectral signatures," *IEEE Geoscience and Remote Sensing Letters* **10**(2), pp. 381–385, 2013.

[50] E. Xing, A. Ng, M. Jordan, and S. Russell, "Distance metric learning, with application to clustering with side-information," *Advances in neural information processing systems* **15**, pp. 505–512, 2002.

[51] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning distance functions using equivalence relations," in *In Proceedings of the Twentieth International Conference on Machine Learning*, 2003.

[52] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon, "Information-theoretic metric learning," in *Proceedings of the 24th international conference on Machine learning*, pp. 209–216, ACM, 2007.

[53] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *Journal of Machine Learning Research* **6**, pp. 1705–1749, 2005.

[54] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, USA: Academic Press, 4th ed., 2008.

[55] G. M. Smith and E. J. Milton, "The use of the empirical line method to calibrate remotely sensed data to reflectance," *International Journal of Remote Sensing* **20**(13), pp. 2653–2662, 1999.

[56] E. Karpouzli and T. Malthus, "The empirical line method for the atmospheric correction of IKONOS imagery," *International Journal of Remote Sensing* **24**(5), pp. 1143–1150, 2003.

[57] W. Lucht, C. Schaaf, and A. Strahler, "An algorithm for the retrieval of albedo from space using semiempirical BRDF models," *Geoscience and Remote Sensing, IEEE Transactions on* **38**, pp. 977–998, Mar 2000.

[58] C. B. Schaaf, F. Gao, A. H. Strahler, W. Lucht, X. Li, T. Tsang, N. C. Strugnell, X. Zhang, Y. Jin, J.-P. Muller, *et al.*, "First operational BRDF, albedo nadir reflectance products from MODIS," *Remote sensing of Environment* **83**(1), pp. 135–148, 2002.

[59] C. Song, C. E. Woodcock, K. C. Seto, M. P. Lenney, and S. A. Macomber, "Classification and change detection using Landsat TM data: When and how to correct atmospheric effects?," *Remote Sensing of Environment* **75**(2), pp. 230–244, 2001.

[60] P. S. Chavez, "Radiometric calibration of Landsat Thematic Mapper multispectral images," *Photogrammetric Engineering and Remote Sensing* **55**(9), pp. 1285–1294, 1989.

[61] T. Chan and L. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on* **10**, pp. 266–277, Feb 2001.

[62] R. Richter, *Atmospheric/Topographic Correction for Satellite Imagery. ATCOR-2/3 User Guide, Version 8.2. (online)*, 2012.

[63] "http://landsat.usgs.gov/landsat8.php."

[64] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *Geoscience and Remote Sensing, IEEE Transactions on* **46**(1), pp. 228–236, 2008.

[65] Y. Zhang, "A new automatic approach for effectively fusing Landsat 7 as well as IKONOS images," in *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International*, **4**, pp. 2429–2431, IEEE, 2002.

[66] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **12**(7), pp. 629–639, 1990.

[67] J. R. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective 2/e*, Pearson Education India, 2009.

[68] F. Kriegler, W. Malila, R. Nalepka, and W. Richardson, "Preprocessing transformations and their effects on multispectral recognition," in *Remote Sensing of Environment, VI*, **1**, p. 97, 1969.

[69] A. K. Shackelford and C. H. Davis, "Fully automated road network extraction from high-resolution satellite multispectral imagery," in *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, **1**, pp. 461–463, IEEE, 2003.

[70] X. Jin and C. H. Davis, "An integrated system for automatic road mapping from high-resolution multi-spectral satellite imagery by information fusion," *Information Fusion* **6**(4), pp. 257–273, 2005.

[71] J. Y. Lee, "Automated extraction of road networks from IKONOS data in urban area," in *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International*, **1**, pp. 4 pp.–, July 2005.

[72] B. Chen, A. Vodacek, and N. Cahill, "Novel spectral similarity measure for high resolution urban scenes," in *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pp. 6637–6640, July 2012.

[73] B. Chen, A. Vodacek, and N. D. Cahill, "A novel adaptive scheme for evaluating spectral similarity in high-resolution urban scenes," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **6**(3), pp. 1376–1385, 2013.

[74] C. Panagiotakis, E. Kokinou, and A. Sarris, "Curvilinear structure enhancement and detection in geophysical images," *Geoscience and Remote Sensing, IEEE Transactions on* **49**(6), pp. 2040–2048, 2011.

[75] M. Haklay and P. Weber, "Openstreetmap: User-generated street maps," *Pervasive Computing, IEEE* **7**(4), pp. 12–18, 2008.

[76] J. Bennett, *OpenStreetMap*, Packt Publishing, 2010.

[77] M. Haklay, "How good is volunteered geographical information? a comparative study of OpenStreetMap and Ordnance Survey datasets," *Environment and planning. B, Planning & design* **37**(4), p. 682, 2010.

[78] R. Canavosio-Zuzelski, P. Agouris, and P. Doucette, "A photogrammetric approach for assessing positional accuracy of OpenStreetMap© roads," *ISPRS International Journal of Geo-Information* **2**(2), pp. 276–301, 2013.

[79] "https://blog.openstreetmap.org/2015/03/12/two-million-contributors/."

[80] "http://blog.osmfoundation.org/2012/03/08/welcome-apple."

[81] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing* **21**(11), pp. 977–1000, 2003.

[82] D. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *Pattern Recognition* **13**(2), pp. 111 – 122, 1981.

[83] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**, pp. 603–619, May 2002.